

RESULT 4
US-09-509-814A-1
Sequence 1, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIRO
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUM
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP99/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 639
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (23)..(23)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (29)..(29)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (32)..(32)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (53)..(53)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (70)..(70)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (74)..(74)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (89)..(89)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (102)..(102)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (105)..(105)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (128)..(128)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (130)..(130)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature

LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (146)..(146)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (160)..(160)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (165)..(165)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (172)..(172)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (183)..(183)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (187)..(187)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (194)..(194)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (286)..(286)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (306)..(306)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (324)..(324)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (369)..(369)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (431)..(431)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (501)..(501)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (531)..(531)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (541)..(541)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (584)..(584)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (591)..(591)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (594)..(594)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (595)..(595)

```
;
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (596)..(596)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (611)..(611)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (632)..(632)
; OTHER INFORMATION: Xaa is any amino acid
;
US-09-509-814A-1

Query Match          95.7%; Score 2150; DB 4; Length 639;
Best Local Similarity 96.1%; Pred. No. 1.8e-164;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSVHEAFRGKITALVALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSVHEAFRGKITALVALGRTN 265

QY 61 NANDNGHGHVAGSVLGNSTNGKNGAPQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
Db 266 NANDNGHGHVAGSVLGNSTNGKNGAPQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 325

QY 121 AGARIHTNSGVAAGVAYTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSGVAAGVAYTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI 385

QY 181 TVGATENLRFPSFGSYADNINHVAFSSRGPTDKGRIPDVAPGTFILSRSSLPDSSF 240
Db 386 TVGATENLRFPSFGSYADNINHVAFSSRGPTDKGRIPDVAPGTFILSRSSLPDSSF 445

QY 241 WANHDSKIAYMGTSMATPIVAGNVAQLREHFVQRGITPKPSLLKALIAAGADIGLY 300
Db 446 WANHDSKIAYMGTSMATPIVAGNVAQLREHFVQRGITPKPSLLKALIAAGADIGLY 505

QY 301 PNGNQGWGRVLDKSLNVAIVNNESSLSQATYSFTATAGKPLKISLWSDAPASTTA 360
Db 506 PNGNQGWGRVLDKSLNVAIVNNESSLSQATYSFTATAGKPLKISLWSDAPASTTA 565

QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPSPYNDNNGRNVNFIINAPQSGTGTIEVQAYN 420
Db 566 SVTLVNDLNLVITAPNGTQVGNDFTPSPYNDNNGRNVNFIINAPQSGTGTIEVQAYN 625

QY 421 VPVGQPTESLAIVN 434
Db 626 VPVGQPTESLAIVN 639

RESULT 5
US-09-509-814A-2
; Sequence 2, Application US/09509814A
; Patent No. 6378227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
```

```
;
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (30)..(30)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (47)..(47)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (48)..(48)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (54)..(54)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (71)..(71)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (75)..(75)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (90)..(90)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (103)..(103)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (106)..(106)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (129)..(129)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (131)..(131)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (132)..(132)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (134)..(134)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (147)..(147)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (149)..(149)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (161)..(161)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (166)..(166)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (173)..(173)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (184)..(184)
; OTHER INFORMATION: Xaa is any amino acid
```

NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (432)..(432)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid
US-09-814A-2

Query Match 95.7%; Score 2150; DB 4; Length 640;
Best Local Similarity 96.1%; Pred. No. 1.8e-164;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAGSSYGLYCGQGVAVADTGLDGRNDSMHEAFRGKITALVALGRTN 60
Db 207 NDVARGIVKADVAGSSYGLYCGQGVAVADTGLDGRNDSMHEAFRGKITALVALGRTN 266

Qy 61 NANTNTHGHTHAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAVS 120
Db 267 NANTNTHGHTHAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAVS 326

Qy 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDWTLFAAGNENGPNGGTTISAPGTAKNAI 180
Db 327 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDWTLFAAGNENXPNGGTTISAPGTAKNAI 385
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDVWAPGTFTLSARSSSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDVWAPGTFTLSARSSSLAPDSSF 446
Qy 241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLY 300
Db 447 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLY 506
Qy 301 PNGQGWGRVTLDSINVAAYVNSSSLSTSQATYSFTATAGPKLSLWSDAPASTTA 360
Db 507 PNGQGWGRVTLDSINVAAYVNSSSLSTSQATYSFTATAGPKLSLWSDAPASTTA 566
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTSFYNDNMDGRNENVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTSFYNDNMDGRNENVFINAPQSGTYTIEVQAYN 626
Qy 421 VPVGPOTFSLAIVN 434
Db 627 VPVGPOTFSLAIVN 640

RESULT 6
US-08-873-479-42
; Sequence 42, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58917010 No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-873-479-42

Query Match 94.4%; Score 2120.5; DB 2; Length 641;
Best Local Similarity 93.3%; Pred. No. 4.2e-162;
Matches 405; Conservative 20; Mismatches 8; Indels 1; Gaps 1;

```
QY 1 NDVARGIVKADVAQSSVGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 209 NDVARGIVKADVAQNFGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 269
QY 61 NANDTNGHGHVAGSVLNGSTKKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 269 NANDPNGHGHVAGSVLNG-ATTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 327
QY 121 AGARIHNSNGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 328 AGARIHNSNGAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 387
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSLAPDSFF 240
DB 388 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSLAPDSFF 447
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGGITPKPSLLKAAIAGAADI GLGY 300
DB 448 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGGITPKPSLLKAAIAGAADVGLGF 507
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 508 PNGNQGWGRVTLDKSLNVAFVNETSPLSTSKATYSFTACAGKPLKISLVMSDAPGSTTA 567
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 420
DB 568 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 627
QY 421 VPVGPTQFSLAIVN 434
DB 628 VPVSPQTFFSLAIVH 641
```

```
RESULT 7
US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Roggen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6303752o No. 6303752disk of No. 6303752th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:
```

```
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: Bacillus sp. Y
US-09-104-623A-4
Query Match 88.2%; Score 1981.5; DB 4; Length 433;
Best Local Similarity 87.1%; Pred. No. 3.6e-151;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSVGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQNFGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTKKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDPNGHGHVAGSVLNG-ALANKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHNSNGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 120 AGARIHNSNGAPVNGAYTTANSQVDEYVRNNDMTVLFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSLAPDSFF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSLAPDSFF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGGITPKPSLLKAAIAGAADI GLGY 300
DB 240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVKNGGITPKPSLLKAAIAGATDVGLGY 299
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYVNEATATGKATYSFQAQAGKPLKISLVMTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 420
DB 360 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 419
QY 421 VPVGPTQFSLAIVN 434
DB 420 VPSGPQFSLAIVH 433
```

RESULT 8

```
US-09-019-532-4
; Sequence 4, Application US/09019532B
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/00015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4
```

```
Query Match 88.2%; Score 1981.5; DB 4; Length 433;
Best Local Similarity 87.1%; Pred. No. 3.6e-151;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;
```


QY 1 NDVARGIVKADVAOSSYGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQNNGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 61 NASDPNGHGHVAGSVLNG -ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
 QY 121 AGARIHTNSWGAAYVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
 Db 120 AGARIHTNSWGAAYVNGAYTANSROVDYVRNNDMTILFAAGNEGPNGGTTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPOGTFILSARSLAPDSSF 240
 Db 180 TVGATENRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAPOGTFILSARSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMAATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAAGAADIGLY 300
 Db 240 WANYNSKYAYMGGTSMAATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAAGAADIGLY 299
 QY 301 PNGNQGWGRVTLDSLVNVAAYNNESSLSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 300 PSGDQGWGRVTLDSLVNVAAYNEATATGKATYSFOAQAGKPLKISLVMTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTYGVNDFTSPYNDWMDGRNENNVFINAPQSGTYTIEVQAYN 420
 Db 360 SYTLVNDLNLVITAPNGTYGVNDFTSPYNDWMDGRNENNVFINAPQSGTYTIEVQAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 Db 420 VPSGPQRFSLAIVH 433

RESULT 9

US-09-338-746-4

; Sequence 4, Application US/09338746

; Patent No. 6638526

; GENERAL INFORMATION:

; APPLICANT: Deussen, Heinz-Josef

; APPLICANT: Olsen, Arne A.

; APPLICANT: Fatum, Tine M.

; APPLICANT: Roggen, Erwin L.

; TITLE OF INVENTION: A Polypeptide-Polymer Conjugate

; FILE REFERENCE: 5619-200-US

; CURRENT APPLICATION NUMBER: US/09/338,746

; CURRENT FILING DATE: 1998-06-23

; EARLIER APPLICATION NUMBER: PA 1998 00809

; EARLIER FILING DATE: 1998-06-23

; EARLIER APPLICATION NUMBER: 60/091,461

; EARLIER FILING DATE: 1998-07-01

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 433

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-338-746-4

Query Match 88.2%; Score 1981.5; DB 4; Length 433;

Best Local Similarity 87.1%; Pred. No. 3.6e-151;

Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAOSSYGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQNNGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 61 NASDPNGHGHVAGSVLNG -ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
 QY 121 AGARIHTNSWGAAYVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180

Db 120 AGARIHTNSWGAAYVNGAYTANSROVDYVRNNDMTILFAAGNEGPNGGTTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPOGTFILSARSLAPDSSF 240
 Db 180 TVGATENRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAPOGTFILSARSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMAATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAAGAADIGLY 300
 Db 240 WANYNSKYAYMGGTSMAATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAAGAADIGLY 299
 QY 301 PNGNQGWGRVTLDSLVNVAAYNNESSLSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 300 PSGDQGWGRVTLDSLVNVAAYNEATATGKATYSFOAQAGKPLKISLVMTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTYGVNDFTSPYNDWMDGRNENNVFINAPQSGTYTIEVQAYN 420
 Db 360 SYTLVNDLNLVITAPNGTYGVNDFTSPYNDWMDGRNENNVFINAPQSGTYTIEVQAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 Db 420 VPSGPQRFSLAIVH 433

RESULT 10

US-08-873-479-43

; Sequence 43, Application US/08873479

; Patent No. 5891701

; GENERAL INFORMATION:

; APPLICANT: Sloma, Alan

; APPLICANT: Lynne, Christianson

; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide

; TITLE OF INVENTION: Having Protease Activity

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5891701 No. 5891701disk of No. 5891701th America

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: Fast-Seq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/873,479

; FILING DATE: 12-JUN-1997

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Agis, Cheryl H

; REGISTRATION NUMBER: 34,086

; REFERENCE/DOCKET NUMBER: 5251.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 635 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-873-479-43

Query Match 88.2%; Score 1981.5; DB 2; Length 635;

Best Local Similarity 87.1%; Pred. No. 6.1e-151;

Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAOSSYGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 Db 203 NDVARGIVKADVAQNNGLYGGQIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 262

Qy	61	NANDTNGHGT	HVAGSVL	NGSTNGK	MAQANL	VFQST	IMDSGG	GLG	PSNL	QTLF	SOAYS	120																																											
		:	:	:	:	:	:	:	:	:	:																																												
Db	263	NASDPNGHGT	HVAGSVLGN	-ALNKG	MAQANL	VFQST	IMDSGG	GLG	PSNL	T	LF	SOAWN	321																																										
		:	:	:	:	:	:	:	:	:	:																																												
Qy	121	AGARHTNS	WGAAV	NGAVY	T	TSRNV	DDVYR	KND	MTIL	FAAG	NEG	PGNGT	II	AS	PGTAK	NAI	180																																						
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																																							
Db	322	AGARHTNS	WGAF	PUNGAY	TANS	RQV	DEYV	RND	MTVL	FAAG	NEG	PGNS	G	II	AS	PGTAK	NAI	381																																					
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																																							
Qy	181	TVGATEN	LR	SP	FGS	YAD	NH	VAQ	FSS	RGP	T	KDGR	IK	P	DY	MA	PGT	FTIL	SAR	SS	LAP	DS	SF	240																															
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																																
Db	382	TVGATENY	R	SP	FGS	IAD	NP	HN	IAQ	FSS	R	GT	R	D	GR	I	K	P	D	V	T	A	P	G	T	FTIL	SAR	SS	LAP	DS	SF	441																							
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																							
Qy	241	WAN	HS	KY	A	Y	M	G	G	T	S	M	A	T	P	I	V	A	G	N	V	A	Q	L	R	E	H	F	V	K	R	G	I	T	P	K	P	S	L	L	K	A	A	L	I	A	G	A	D	T	G	L	G	Y	300
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:						
Db	442	WANTNS	KY	A	Y	M	G	G	T	S	M	A	T	P	I	V	A	G	N	V	A	Q	L	R	E	H	F	I	K	R	G	I	T	P	K	P	S	L	L	K	A	A	L	I	A	G	A	D	T	G	L	G	Y	501	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:						
Qy	301	PNG	Q	G	W	G	R	V	T	L	D	S	L	N	V	A	Y	N	E	S	S	L	S	T	S	O	K	A	T	Y	S	T	A	T	A	G	R	P	L	K	I	S	L	V	M	S	D	A	P	A	S	T	T	A	360
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:						
Db	502	PSG	D	Q	G	W	G	R	V	T	L	D	S	L	N	V	A	Y	N	E	A	L	T	A	G	K	A	T	Y	S	F	Q	A	G	R	P	L	K	I	S	L	V	M	D	A	P	S	T	T	A	561				
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:						
Qy	361	S	V	T	L	N	D	L	N	L	V	I	T	A	P	N	G	T	S	P	Y	N	D	P	T	S	P	Y	N	D	G	R	N	V	N	F	I	N	A	P	S	G	T	Y	T	I	E	V	Q	A	N	420			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:						
Db	562	S	T	L	N	D	L	O	L	V	I	T	A	P	N	G	K	Y	V	G	N	D	S	F	Y	P	D	N	N	D	G	R	N	V	N	F	I																		

```

RESULT 11
US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 6555355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349, 204-US
; CURRENT APPLICATION NUMBER: US/09/512, 251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-512-251A-10

```

Query Match	70.4%;	Score 1581.5;	DB 4;	Length 345;
Best Local Similarity	95.3%;	Pred. No. 3.4e-119;		
Matches 303;	Conservative 10;	Mismatches 4;	Indels 1;	Gaps 1
Qy	1	NDVARGIKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTN	60	
Db	29	NDVARGIKADVAQNFGYLGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTN	88	
Qy	61	NANDTNGHGTTHVAGSVLGNSTGNGKMAQANLVFQSIINDSGGGJGLHPSMLQTLFQAYS	120	
Db	89	NANDPNHGHTHVAGSVLGN-A'NKGMAQANLVFQSIINDSGGGJGLHPSMLQTLFQAYS	147	
Qy	121	AGARIHTNSGWAIVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGTT-SAPGTAKNAI	180	
Db	148	AGARIHTNSGAPVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGTT-SAPGTAKNAI	207	
Qy	181	TVGATENLPPSGSYADNINHVAQSSRGPKDGRHKPDVMAQPTFILSARSLAPDSSF	240	
Db	208	TVGATENLPPSGSYADNINHVAQSSRGPKDGRHKPDVMAQPTYILSARSLAPDSSF	267	
Qy	241	WANHDSKYAWMGTSMATPIVAGNVAQLREHFVKNRGTTPKPSLLKAALTAGAADILGLY	300	
Db	268	WANHDSKYAWMGTSMATPIVAGNVAQLREHFVKNRGVTPKPSLLKAALTAGAADVLGF	337	

```

Qy      301  PGNQOQGRVTLDKSLNV 318
        |||||
        |||||
Db      328  PGNQOQGRVTLDKSLNV 345
        |||||
        |||||

RESULT 12
US-09-515-150A-10
; Sequence No. Application US/09515150A
; Patent No. 6559938
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variant
; FILE REFERENCE: 5348.204-US
; CURRENT APPLICATION NUMBER: US/09/515
; CURRENT FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-515-150A-10

```

Query Match	70.4%; Score
Best Local Similarity	95.3%; Pred.
Matches	303; Conservative
Qy	1 NDVARGIVKADVAQSSVGLYGQGGI
Db	29 NDVARGIVKADVAQNFGSLYGQGGI
Qy	61 NANDTNGHGTHVAGSVLGNQSTNGK
Db	89 NANDPNGHGTHVAGSVLGN-ATNKG
Qy	121 AGARIHTNSWGAAVNGAYTTTDSRNV
Db	148 AGARIHTNSWGAPVNGAYTTTDSRNV
Qy	181 TVGATENLRSFGSYADININHVAGQF
Db	208 TVGATENLRSFGSYADININHVAGQF
Qy	241 WANHDSKYAMGCGTSMATPIVAGNV
Db	268 WANHDSKYAMGCGTSMATPIVAGNV
Qy	301 PNGNQGWGRVTLDKSLNV 318
Db	328 PNGNQGWGRVTLDKSLNV 345
RESULT 13	
US-09-196-281-13	
; Sequence 13 Application US/09196281A	
; Patent No. 6603458	
; GENERAL INFORMATION:	
; APPLICANT: Hansen, Peter K.	
; APPLICANT: Bauditt, Peter	
; APPLICANT: Mikkelson, Frank	
; TITLE OF INVENTION: Protease Variant	
; FILE REFERENCE: 5435,200-US	
; CURRENT APPLICATION NUMBER: US/09/19	
; CURRENT FILING DATE: 1998-11-19	
; EARLIER APPLICATION NUMBER: 1332/97	
; EARLIER FILING DATE: 1997-11-21	
; NUMBER OF SEQ ID NOS: 18	
; SOFTWARE: FastSeq for Windows Version	
; SEQ ID NO 13	

```

; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-196-281-13

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 3.4e-119;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVIVAVADTGLDTRGNDSSMHEAPRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAGQNNFGLYGGQGVIVAVADTGLDTRGNDSSMHEAPRGKITALYALGRTN 88
QY 61 NANDTNHGHTHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGSLPNLQTLFSAQYS 120
Db 89 NANDPENGHTHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGSLPANLQTLFSAQYS 147
QY 121 AGARLHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
Db 148 AGARLHTNSGAPVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGSGTISAPGTAKNAI 207
QY 181 TVGATENLRPFSGYADNINHAQVFSRGTQDGR.KPDVMAFGTIFLSARSSLAPDSSF 240
Db 208 TVGATENLRPFSGYADNINHAQVFSRGTQDGR.KPDVMAFGTYILSARSSLAPDSSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAOLREHFVKNRGITPKPSLLKAAIIAGAADIGLGY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAOLREHFVKNRGVTPKPSLLKAAIIAGAADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 14
US-08-894-818B-1
; Sequence 1, Application US/088994818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 233285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:54:44 ; Search time 40 Seconds
(without alignments)
2291.013 Million cell updates/sec

Title: US-09-985-689A-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVPGPQFFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2242	99.8	434	10	US-09-985-689A-1
2	2242	99.8	434	15	US-10-385-662-2
3	2186	97.3	434	10	US-09-985-689A-2
4	2138	95.1	434	10	US-09-985-689A-6
5	2120.5	94.4	433	10	US-09-985-689A-7
6	1993.5	88.7	433	10	US-09-985-689A-5
7	1989.5	88.5	433	10	US-09-985-689A-3
8	1982.5	88.2	433	10	US-09-985-689A-4
9	1581.5	70.4	345	14	US-10-336-324-10
10	1581.5	70.4	345	14	US-10-403-105-13
11	447.5	19.9	659	13	US-10-090-624-12
12	409	18.2	412	13	US-10-090-624-1
13	409	18.2	522	13	US-10-090-624-4
14	409	18.2	654	13	US-10-090-624-16
15	366	16.3	1079	14	US-10-112-488-39

16	363.5	16.2	1208	14	US-10-156-761-13251	Sequence 13251, A
17	348	15.5	1139	14	US-10-156-761-10856	Sequence 10856, A
18	336	15.0	1237	14	US-10-314-857-4	Sequence 4, Appli
19	306.5	13.6	519	15	US-10-084-846A-114	Sequence 114, App
20	306.5	13.6	19725	15	US-10-084-846A-4	Sequence 4, Appli
21	305.5	13.6	1398	13	US-10-090-624-6	Sequence 6, Appli
22	305.5	13.6	1398	13	US-10-090-624-6	Sequence 6, Appli
23	280.5	12.5	595	10	US-09-927-827-59	Sequence 59, Appl
24	278	12.4	580	10	US-09-927-827-55	Sequence 55, Appl
25	276	12.3	418	9	US-09-966-921A-2	Sequence 2, Appli
26	275	12.1	1101	14	US-10-156-761-12934	Sequence 12934, A
27	251	11.2	280	14	US-10-209-812-2	Sequence 2, Appli
28	251	11.2	397	10	US-09-779-334A-5	Sequence 5, Appli
29	246.5	11.0	271	10	US-09-813-408-2	Sequence 2, Appli
30	246.5	11.0	379	10	US-09-813-408-6	Sequence 6, Appli
31	244	10.9	271	14	US-10-242-549-56	Sequence 56, Appl
32	244	10.9	271	14	US-10-242-549-46	Sequence 46, Appl
33	244	10.9	271	14	US-10-242-549-48	Sequence 48, Appl
34	244	10.9	271	14	US-10-242-549-50	Sequence 50, Appl
35	244	10.9	271	14	US-10-242-549-52	Sequence 52, Appl
36	244	10.9	271	14	US-10-242-549-54	Sequence 54, Appl
37	242.5	10.8	271	14	US-10-242-549-60	Sequence 60, Appl
38	242	10.8	271	14	US-09-927-827-60	Sequence 60, Appl
39	242	10.8	271	14	US-10-242-549-44	Sequence 44, Appl
40	240	10.7	269	8	US-08-322-678-10	Sequence 10, Appl
41	240	10.7	269	9	US-09-837-235-16	Sequence 16, Appl
42	240	10.7	269	9	US-09-060-854B-6	Sequence 6, Appli
43	240	10.7	269	9	US-09-975-139-1	Sequence 1, Appli
44	240	10.7	269	9	US-09-976-414-8	Sequence 8, Appli
45	240	10.7	269	10	US-09-736-116-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1

US-09-985-689A-1
; Sequence 1, Application US/09985689A

; Publication NO. US20030022351A1

; GENERAL INFORMATION:

; APPLICANT: HATADA, YUJI

; APPLICANT: OGAWA, AKINORI

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SATO, TSUYOSHI

; APPLICANT: ARAKI, HIROYUKI

; APPLICANT: SUMITOMO, NOBUYUKI

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; TITLE OF INVENTION: Alkaline proteases

; FILE REFERENCE: 215483US0

; CURRENT APPLICATION NUMBER: US/09/985,689A

; PRIOR FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: JP P2000-355166

; PRIOR FILING DATE: 2000-11-22

; PRIOR APPLICATION NUMBER: JP P2001-114048

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 434

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-985-689A-1

Query Match 99.8%; Score 2242; DB 10; Length 434;
Best Local Similarity 99.8%; Pred. No. 3.9e-187; Indels 0; Gaps 0;
Matches 433; Conservative 1; Mismatches 0;

QY 1 NDVARGIVKADVAQSSVGLYGGQGIIVAVADTGLDGRNDSNMHEAFRCKITAIYALGRTN 60

Db 1 NDVARGIVKADVAQSSVGLYGGQGIIVAVADTGLDGRNDSNMHEAFRCKITAIYALGRTN 60

QY 61 NANDTINGHGTAVGSLVNGSTNKMAQANLVFQSIMDSGGGLGGLPNSLTLPQSQAYS 120

```

Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPNNLTLSQAYS 120
      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240
      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240
      241  WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
      241  WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
      301  PNGNQGRVTLTKSLNVAVYNESSSLSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
      301  PNGNQGRVTLTKSLNVAVYNESSSLSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
      361  SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
      361  SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
      421  VPVGPQTFSLAIVN 434
      421  VPVGPQTFSLAIVN 434

```

RESULT 2

```

US-10-385-662-2
; Sequence 2, Application US/10385662
; Publication No. US2004002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10385,662
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp. KSM-KP43
US-10-385-662-2

```

```

Query Match      99.8%; Score 2242; DB 15; Length 434;
Best Local Similarity 99.8%; Pred. No. 3.9e-187;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1  NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Qy      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPNNLTLSQAYS 120
Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPNNLTLSQAYS 120

```

```

Qy      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Qy      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240
Db      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240
Qy      241  WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
Db      241  WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
Qy      301  PNGNQGRVTLTKSLNVAVYNESSSLSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
Db      301  PNGNQGRVTLTKSLNVAVYNESSSLSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
Qy      361  SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
Db      361  SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
Qy      421  VPVGPQTFSLAIVN 434
Db      421  VPVGPQTFSLAIVN 434

```

RESULT 3

```

US-09-985-689A-2
; Sequence 2, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-2

```

```

Query Match      97.3%; Score 2186; DB 10; Length 434;
Best Local Similarity 96.3%; Pred. No. 3e-182;
Matches 418; Conservative 14; Mismatches 2; Indels 0; Gaps 0;

Qy      1  NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Qy      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPNNLTLSQAYS 120
Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPNNLTLSQAYS 120
Qy      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Qy      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240
Db      181  TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFTILSARSSLAPDSSF 240

```

```
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
Db 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADVGLGY 300
QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGVNDFTSPYNDNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 420
Db 361 SVTLVNDLNLVITAPNGTQVGVNDFTSPYNDNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 4
US-09-985-689A-6
; Sequence 6, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-6

Query Match 95.1%; Score 2138; DB 10; Length 434;
Best Local Similarity 93.3%; Pred. No. 4.7e-178;
Matches 405; Conservative 20; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRGNDSSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRGNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 121 AGARIHTNSWGAPVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKGRIPKPDVMAPTGTFILSARSLAPDSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKGRIPKPDVMAPTGTFILSARSLAPDSF 240
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
Db 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434
```

```
Db 301 PSNGQGWGRVTLDKSLNVAFVNETSSLSSTQKATYSFTAQSGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGVNDFTSPYNDNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 420
Db 361 SVTLVNDLNLVITAPNGTQVGVNDFTAPYDNNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 5
US-09-985-689A-7
; Sequence 7, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-7

Query Match 94.4%; Score 2120.5; DB 10; Length 433;
Best Local Similarity 93.3%; Pred. No. 1.6e-176;
Matches 405; Conservative 20; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRGNDSSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSNFGLYGQGVAVADTGLDTRGNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDTNGHGHVAGSVLNG-ATNKGMAPOANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 119
QY 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAPVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPGSGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKGRIPKPDVMAPTGTFILSARSLAPDSF 240
Db 180 TVGATENLRPSFGSYADNINHVAQFSRSGPTKGRIPKPDVMAPTGTFILSARSLAPDSF 239
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
Db 240 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADVGLGF 299
QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 300 PNGNQGWGRVTLDKSLNVAFVNETSPLSTQKATYSFTAQAGKPLKISLVWSDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVGVNDFTSPYNDNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 420
Db 360 SCTLVNDLNLVITAPNGTQVGVNDFTAPYDNNWDGRNNVNFVFINAPQSGTYYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
```

```
Db 420 VPVSPQTFSLAIHV 433
|||||
RESULT 6
US-09-985-689A-5
; Sequence 5, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-5

Query Match 88.7%; Score 1993.5; DB 10; Length 433;
Best Local Similarity 87.6%; Pred. No. 1.9e-165;
Matches 380; Conservative 29; Mismatches 24; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHAEFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHAEFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLSQAYS 120
Db 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLSQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTLFAAGNEGPNNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTLFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNHNHVAQSSRGTQDGRKPDVMAFGTFLSARSSSLAPDSSF 240
Db 180 TVGATENYRPSFGSLADNPNHIAQSSRGATRDGRKPDVTAPGTFLSARSSSLAPDSSF 239
QY 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKRGITPKPSLLKAAALIAAGADIGLY 300
Db 240 WANYNSKYAMGTSMTATPIVAGNVAQLREHFVKRGITPKPSLLKAAALIAAGADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAYNVESLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 300 PSGDQGWGRVTLDKSLNVAYNVEATATGQKATYSFOAQAGKPLKISLWTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPSYDNNWDRNNVNFVINAPOSQGTITIEVQAYN 420
Db 360 SVTLVNDLNLVITAPNGQKIVGNDFSYDNNWDRNNVNFVINAPOSQGTITIEVQAYN 419
QY 421 VPVGPQTFSLAIYN 434
Db 420 VPSPQRFSLAIHV 433

RESULT 7
US-09-985-689A-3
; Sequence 3, Application US/09985689A
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-3

Query Match 88.5%; Score 1989.5; DB 10; Length 433;
Best Local Similarity 87.3%; Pred. No. 4.3e-165;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHAEFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHAEFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLSQAYS 120
Db 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLSQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTLFAAGNEGPNNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTLFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNHNHVAQSSRGTQDGRKPDVMAFGTFLSARSSSLAPDSSF 240
Db 180 TVGATENYRPSFGSIADNPNHIAQSSRGATRDGRKPDVTAPGTFLSARSSSLAPDSSF 239
QY 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKRGITPKPSLLKAAALIAAGADIGLY 300
Db 240 WANYNSKYAMGTSMTATPIVAGNVAQLREHFVKRGITPKPSLLKAAALIAAGADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAYNVESLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 300 PSGDQGWGRVTLDKSLNVAYNVEATALTGQKATYSFQAGKPLKISLWTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPSYDNNWDRNNVNFVINAPOSQGTITIEVQAYN 420
Db 360 SVTLVNDLNLVITAPNGQKIVGNDFSYDNNWDRNNVNFVINAPOSQGTITIEVQAYN 419
QY 421 VPVGPQTFSLAIYN 434
Db 420 VPSPQRFSLAIHV 433

RESULT 8
US-09-985-689A-4
; Sequence 4, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
```



```
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985,689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-4

Query Match      88.2%; Score 1982.5; DB 10; Length 433;
Best Local Similarity 87.1%; Pred. No. 1.7e-164;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEVVRNDMTVLFAAGNEGPSNGTISAPGTAKNAI 179
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAPTILSARSSLAPDSF 240
Db 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATGRIPKPDVAPGTFTILSARSSLAPDSF 239
Qy 241 WANHDSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 240 WANYSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAATDVLGY 299
Qy 301 PNGQGWGRVTLKSLNVAVNVSLSSTQKATYSFTAGKPLKISLWSDAPASTTA 360
Db 300 PNGQGWGRVTLKSLNVAVNVEATATGKATYSFQAQAGKPLKISLWTDAPGSTTA 359
Qy 361 SVTLVNDLNLVITAPNGTQVGVNDFTSPYDNDGRNVDNENFINAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLNLVITAPNGKQVGVNDFTSPYDNDGRNVDNENFINAPQSGTYTIEVQAYN 419
Qy 421 VPGQPQTFSLAIVN 434
Db 420 VPSGPQRFSLAIVH 433

RESULT 9
US-10-336-324-10
; Sequence 10, Application US/10336324
; Publication No. US20030176304A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/10/336,324
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US/09/512,251A
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-336-324-10

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 1.2e-129;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPENGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 147
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPSNGTISAPGTAKNAI 207
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAPTILSARSSLAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAPTILSARSSLAPDSF 267
Qy 241 WANHDSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 268 WANHDSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
Qy 301 PNGQGWGRVTLKSLNV 318
Db 328 PNGQGWGRVTLKSLNV 345

RESULT 10
US-10-403-105-13
; Sequence 13, Application US/10403105
; Publication No. US20030180933A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/10/403,105
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US/09/196,281A
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1332/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-403-105-13

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 1.2e-129;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPENGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 147
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPSNGTISAPGTAKNAI 207
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAPTILSARSSLAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAPTILSARSSLAPDSF 267
Qy 241 WANHDSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 268 WANHDSKYAYVGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
Qy 301 PNGQGWGRVTLKSLNV 318
Db 328 PNGQGWGRVTLKSLNV 345
```

Db 148 AGARIHTNSWGPVANGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSGTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSADYNHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSSSLAPDSSF 240
Db 208 TVGATENLRPSFGSADYNHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPRLSLKAALIAAGADIGLY 300
Db 268 WANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPRLSLKAALIAAGADIGLY 327
QY 301 PNGNQGWRVTLDKSLNV 318
Db 328 PNGNQGWRVTLDKSLNV 345

RESULT 11
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match 19.9%; Score 447.5; DB 13; Length 659;
Best Local Similarity 29.8%; Pred. No. 3e-30;
Matches 137; Conservative 68; Mismatches 153; Indels 101; Gaps 18;

QY 8 VKADVAQSSVGLYGOGIVAVADTGLDTCGRNDSMHEAFRGKITALLY-ALGRTNNANDTN 66
Db 145 IGADTVNLSGLVDSGVVVAIVDTGIDAN-----HPDLKGVIGWYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNK---GMAPQANLVFOSIM--DSGGGLGLPSNLQTLFSAQYSA 121
Db 199 GHGTHVAGSVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVSTIIAGVDVWVQNKDKY 258
QY 122 GARI-----HTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGT 169
Db 259 GIRVNLISGSSQSDGTLSQAVNNAWA-----GIVCVVAGNSGFNYT 306
QY 170 ISAPGTAKNAITVGATENLRPSFGSADYNHVAQFSSRGPTKDGRIKPDVMAPTGFIIS 229
Db 307 VGSRAAASKVITVGA-----VDGNDNIAFSRSGPTADGRLKPEVVAQVVDIIA 355
QY 230 ARSSLAPDSSFWANHDSKYAYMGSTSMATPIVAG-NVAQLREHFVKNRGITPK--PSLLK 286
Db 356 PRAS-----GTSMGTPINDYITKASGTSMATPHVSGVGLITLOAH-----PSWTFDKVY 404
QY 287 AALIAGA-----ADIGLYPENGQWGRVTLDKSL---NVAYVNESSSLSTSQKATY 335
Db 405 TALIEADIVAPKEIADIAVGA-----GRNVYKAIKYDDYAKLITFGSVADKGSATH 457
QY 336 SFTATAGKELKSLVWSADAPASTASVTLVNDLNLVITAPNGTQVYVGNDFTSYNDNDWG 395
Db 458 TFDVSGATTVTATLYWD-----TGSSDIDLYLPNGNE-VDYSYATAY-----500

QY 396 RNNVENVFINAPSGTYTIEVOAYNVVPVGPQTFSLAIVN 434
Db 501 --CFEKVGYNPTAGTWKVVSYK---GAANYQVDVVS 534

RESULT 12
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match 18.2%; Score 409; DB 13; Length 412;
Best Local Similarity 29.8%; Pred. No. 3.5e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGQGOIVAVADTGLDTCGRNDSMHEAFRGKITALLYALGRTNNAN-----DTNGHGTH 71
Db 22 GYDGSGITIGIITDID-----ASHPDLQGGV-----IGWDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGWAPQANLVFOSIM--DSGGGLGLPSNLQTLFSAQYSAQARI 125
Db 71 VASIAAGTGAASNGKYKGWAPGAKAGIKVLAGDGGSGISTIKGVEMAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAP 173
Db 131 INLSGSSQSDGTALSQAVNNAWA-----GLVWVAAGNSGPNKYTIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSADYNHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSS 233
Db 179 AAASKVITVGA-----VDKYDVTITSFSSRGPTADGRLKPEVVAQVWIIAARAS 227
QY 234 LAPDSSFWANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPK--PSLLKAALIA 291
Db 228 ---GTSWGPINDYITAAAGTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 277
QY 292 GA-----ADIGLYPENGQWGRVTLDKSLNVAIVNESSSLTSOKA-----TYSFT 338
Db 278 TADIVKPEADIADIAVGA-----GRVNAVYKAIN--YDNYAKLVFTGYVANKGSQTHQV 328
QY 339 ATAGKPKLISLVWSADAPASTASVTLVNDLNLVITAPNGTQVYVGNDFTSYNDNDWGRNN 398
Db 329 ISGASFVATLYWDNAN-----SDDLILYLPNGNQ-VDYSYATAY-----G 369
QY 399 VENVFINAPSGTYTIEVOAYNVVPVGPQTFSLAIVN 434
Db 370 FEKVGYNPTDGTWTKVWSYS---GSANYQVDVVS 402

RESULT 13
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Pyrococcus furiosus
NAME/KEY: misc feature
LOCATION: (428)..(428)
OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match 18.2%; Score 409; DB 13; Length 522;
Best Local Similarity 29.6%; Pred. No. 4.9e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGTH 71
DB 22 GYDGSGITIGIDGID-----ASHPDLQKV-----IGWVDFVNGRSYFYDDHGHGTH 70
QY 72 VAGSVLNGSTN-----KGMAPQANLVFQSIW--DSGGGLGLPLSNLQTLFSAQYASAGARI 125
DB 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYGIKV 130
QY 126 HTNSGWA-----AVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGTISAP 173
DB 131 INLSGSSQSDGTDALSQAVNAAWDA-----GLVVVAAGNSGPNKYTIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADINHVAFSSRGPTKDGRIKPDVMAPTGTFILSARSS 233
DB 179 AAASKVITVGA-----VDKYDVITSFSSRGPTADGRUKPEVVAPGNWIIAARAS 227
QY 234 LAPDSSFWANHDSKYAYMGSTMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAALIA 291
DB 228 ---GTSMGQPINDDYTAAPGTSMATPHVAGIAALLQ-----AHPSWTPDKVKLTALIE 277
QY 292 GA-----ADIGLYPENGCGWRVTLKSLNVAYVNESSLSLSOKA-----TYSFT 338
DB 278 TADIVKPEDEIADIAVGA-----GRVNAVYKAIN--YDNVAKLVFTGYVANKGSOHQFV 328
QY 339 ATAGKPLKISLWSDAPASTTASVTLVNDLNLVITAPNGTYQVGNDFTSFYNDNWDGRNN 398
DB 329 ISGASFVATLYWDNAN-----SDLDLYDPNGNQ--VDYSYTAYY-----G 369
QY 399 VENVFIPAPQSGTYTIEVQAYNVPGQTFSLAIWN 434
DB 370 FEKVGYNPTDGTWTIKVVSYS---GSANYQVDVVS 402

RESULT 14
US-10-090-624-16
Sequence 16, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-10-090-624-16
Query Match 18.2%; Score 409; DB 13; Length 654;
Best Local Similarity 29.6%; Pred. No. 6.9e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGTH 71
DB 154 GYDGSGITIGIDGID-----ASHPDLQKV-----IGWVDFVNGRSYFYDDHGHGTH 202
QY 72 VAGSVLNGSTN-----KGMAPQANLVFQSIW--DSGGGLGLPLSNLQTLFSAQYASAGARI 125
DB 203 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYGIKV 262
QY 126 HTNSGWA-----AVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGTISAP 173
DB 263 INLSGSSQSDGTDALSQAVNAAWDA-----GLVVVAAGNSGPNKYTIGSP 310
QY 174 GTAKNAITVGATENLRPSFGSYADINHVAFSSRGPTKDGRIKPDVMAPTGTFILSARSS 233
DB 311 AAASKVITVGA-----VDKYDVITSFSSRGPTADGRUKPEVVAPGNWIIAARAS 359
QY 234 LAPDSSFWANHDSKYAYMGSTMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAALIA 291
DB 360 ---GTSMGQPINDDYTAAPGTSMATPHVAGIAALLQ-----AHPSWTPDKVKLTALIE 409
QY 292 GA-----ADIGLYPENGCGWRVTLKSLNVAYVNESSLSLSOKA-----TYSFT 338
DB 410 TADIVKPEDEIADIAVGA-----GRVNAVYKAIN--YDNVAKLVFTGYVANKGSOHQFV 460
QY 339 ATAGKPLKISLWSDAPASTTASVTLVNDLNLVITAPNGTYQVGNDFTSFYNDNWDGRNN 398
DB 461 ISGASFVATLYWDNAN-----SDLDLYDPNGNQ--VDYSYTAYY-----G 501
QY 399 VENVFIPAPQSGTYTIEVQAYNVPGQTFSLAIWN 434
DB 502 FEKVGYNPTDGTWTIKVVSYS---GSANYQVDVVS 534

RESULT 15
US-10-112-488-39
Sequence 39, Application US/10112488
Publication No. US20030082746A1
GENERAL INFORMATION:
APPLICANT: KIKUCHI, Yoshimi
APPLICANT: DATE, Masayo
APPLICANT: UMEZAWA, Yukiko
APPLICANT: YOKOYAMA, Keiichi
APPLICANT: MATSUI, Hiroshi
TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
FILE REFERENCE: 219286USOCONT
CURRENT APPLICATION NUMBER: US/10/112,488
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: PCT/JP00/06780
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: JP2000-280098
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: JP11-280098
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 70

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 1079
; TYPE: PR1
; ORGANISM: Streptomyces albobogriseolus
US-10-112-488-39

Query Match
Best Local Similarity 16.3%; Score 366; DB 14; Length 1079;
Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

QY 3 VAR-----GIVKADVAQS-----SYGLYQGOQIVAVADTGLDTRNDSSMHEAFRG 48
DB 160 VARVWLDGVRKASLDTSVGGTGPKAMEAGYDGKGVKIAVLDTGVD-----ATHPOLKG 213
QY 49 KITALYALGRINNNANDINGCHTHVAGSVLNGS-----TNKGVAPOANLVFOSIMDSGGGL 104
DB 214 QVTASKNFTSAPTTGDDVVVGHGTHVASTAAGTGAQSKGTGKGVAPAKILNKVKLDDAG-- 271
QY 105 GGLPENLOTLFSQAYSAGARIHTNSMGAANGVAYTTDSRNVDYDYRK--NDWTILF--AA 160
DB 272 FGDDSGILAGMEWAAAQAGADIVNMSLG-----GMDTPTDPLEAAVDKLSAEKGILFAIAA 327
QY 161 GNEGPNGGTISAPGTAKNATVGTATENLRSPFSYADNINHVAQSPSSRGP-TKDGRIKPD 219
DB 328 GNEGQQ--SISGSPGASALTVG-----VDDKKDLADFSSTPRLGJGAVKPD 374
QY 220 VNAPGTFTLSARSSLAPDPSFWANHDSKYAYMGGTSMATPIVAGNVAOLRHFVKNGRIT 279
DB 375 LTAPGVDITAAASAKNGDIAKEVGKPGAGYWTISGTSMATPHVAGAAALLKQHP----- 429
QY 280 KPESLLKAALTAGAADTGLG-YPNGNGOGRVTLDKLSLVAYVNSSSL----- 328
DB 430 WKYAEKKGALTASTKD--GKYTFEFGSGRQVQDKAITQTVIABFVSLSPFGVQOWPHAD 486
QY 329 ---TSQKATYSTATAGKPLKISLWMSD-----APAS---TTASVTLVNLDNLVITAP--NG 377
DB 487 DKPVTIKLTYNLGTEDVTLKLTSTATGPKGKAAPAGFFTLGASTL-----TVPANG 538
QY 378 TQYVGNDFTSYNDWNGRNNVENVFINAPQS-----GTYTIEVQAINV 421
DB 539 TASVDVTADTBLGAVDGTYSAYVATGAQGSVRTAAAVEREVESYNV 586

```

Search completed: March 10, 2004, 15:00:19
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:51:09 ; Search time 20 Seconds
(without alignments)
2087.356 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNVPGVPTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78.*

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519.5	23.1	1743	2 T18279	multidrug resistan
2	492	21.9	1905	2 T18267	multidrug resistan
3	347.5	15.5	444	2 B83891	intracellular alka
4	320.5	14.3	442	2 A69887	intracellular alka
5	311.5	13.9	806	2 A41341	microbial serine p
6	303.5	13.5	1398	2 T28159	pyrolysin (EC 3.4
7	281	12.5	419	1 S25835	subtilisin-type pr
8	280	12.5	799	2 G83753	subtilisin (EC 3.4
9	279.5	12.4	1345	2 T29090	surface layer-asso
10	278	12.4	580	2 S11890	serine proteinase
11	277	12.3	420	1 S23407	subtilisin (EC 3.4
12	265	11.8	715	2 JC4908	alkaline serine pr
13	264	11.7	1331	2 A72647	probable surface l
14	262.5	11.7	513	1 A35742	aqualysin (EC 3.4
15	256.5	11.4	757	2 C84120	subtilisin-type pr
16	253.5	11.3	894	2 F69730	cell wall-associat
17	251	11.2	627	2 D75333	serine proteinase
18	246.5	11.0	402	1 JU0332	alkaline proteinase
19	244.5	10.9	534	1 US0173	alkaline proteinase
20	242.5	10.8	401	2 I39974	serine proteinase
21	240	10.7	380	2 A49778	high-alkaline seri
22	240	10.7	488	2 A11930	proteinase (import
23	238	10.6	519	2 S71451	halolysin R4 (EC 3
24	235.5	10.5	382	1 SUBSN	subtilisin (EC 3.4
25	235	10.5	378	2 A33973	high-alkaline seri
26	235	10.5	382	2 I39780	subtilisin (EC 3.4
27	234	10.4	910	2 C69456	subtilisin sendai
28	233	10.4	1374	2 D72593	hypothetical prote
29	231	10.3	321	1 S27501	alkaline proteinase

30 230.5 10.3 379 1 SUBSCL
31 230.5 10.3 525 2 G84406
32 230.5 10.3 601 2 JC4576
33 226.5 10.1 1052 2 T17093
34 226.5 10.1 1118 2 H97298
35 225.5 10.0 381 2 JH0778
36 225.5 10.0 613 2 S75976
37 225.5 10.0 1167 1 A35086
38 224.5 10.0 769 2 D86335
39 223.5 9.9 381 1 SUBSS
40 223.5 9.9 381 1 SUBSI
41 223.5 9.9 381 2 JQ1487
42 222.5 9.9 275 2 JC1085
43 222 9.9 384 2 JC4802
44 221.5 9.9 272 2 A23624
45 221 1036 2 JC5568

ALIGNMENTS

RESULT 1

T18279
Multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18279
R:Shaulsky, G.; Loomis, W.F.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z18855
A:Accession: T18279
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1743 <SHA>
A:Cross-references: EMBL:U60086; NID:g1399914; PID:g1399915; PID:AA03331.1
C:Genetics:
A:Gene: tagC

Query Match 23.1%; Score 519.5; DB 2; Length 1743;
Best Local Similarity 28.4%; Pred. No. 3.8e+25;
Matches 166; Conservative 78; Mismatches 158; Indels 183; Gaps 24;

QY 19 LYGQOIIVADTGLDTCR---NDS-----SMHEAFRGKITALYALGRNNDANDNGH 68
DB 314 LRKGQILSIADTGLDGSCHFCFSDSKYPIPLNSVNLNHR-KVVTYITTSDDSDKVDGH 372
QY 69 GTHVAGSVLG-----NGSTNKGMAPQANLVFQSIMDSGGGLGL--PSNLQTLFQAY 119
DB 373 GTHICSGAAGTPEDSSVNISSFGSLATDAKAF---FDLASGSSSLTPPSDLKQLYQPLY 429
QY 120 SAGARIHTNSGA-----AVNGAYTDSNRVDYVRKN-DMTILPAAGNEGNGGTIS--A 172
DB 430 DAGARVHCDWSGVSVEGYTGSYSSTASIDDFLTHPDFILRAAGN---NEOYLSLLT 486
QY 173 PGTAKNAITVGATENLR-----PSFGSYADNT----- 199
DB 487 QSTAKNVITVGAHQTHENYITDGNVINYOSSVDINOELICDFDSRYCNYTTAQCCLES 546
QY 200 -----NHVAQFSSRGPTKDGRIKPDVMAPTGFTIL 228
DB 547 NATTGLASCCPTLLRKSVIDAANTOPLLYNENNICSFSKGPETHDRMKPALVAPGEYIT 606
QY 229 SARSSLA-----PDSSFWANHDSKYAWGGTSGTATPIVAGNVAQLREH-----F 272
DB 607 SANSNGANTTQCGGSL-PMTNALLA-IGSTWATSFAPAAATILRLYLVDGYPTGSI 664
QY 273 VKNRGITPKPSLLKAALIAGA-----ADIGLYPNGN-----QGWGRVT 311
DB 665 VESNKLPQTGSLIKALMINNAQLLNGTFLQITSSSITYPSNQVFENFAGALVQGWGAI 724
QY 312 LDKSLNVAYNRES-----SISTSQKATYFT-- 338
DB 725 MSNWLHLVNNNNNNNNKNTSDGITKFDGIGGLDLRLVKPNQWKBESUSTGNTSYCYTK 784

```

339 -----ATAGKPLK---ISLWSDAPASTTASVTLVN---DLNLVITAPNG-TQYVGNDFTS 387
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
785 PSSSSSSGNNIPRVVATLVTDTPSYAGAKENLVNLDLTWYYRDNGSIFYSNQGS 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
388 PY---NDWNGRNVNVEINAPQSGTYTIEVQANVPVGPOTFS 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
845 SFLGLAPTQDLNNVEGIVHNPTMTYRFVWAGTNVMPGPNFS 889
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
T18267
multidrug resistance protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T18267
R:Shaulsky, G.; Kuspa, A.; Loomis, W.F.
A:Description: An MDR transporter/sarime protease gene is required for prestalk speciali
A:Reference number: Z18850
A:Accession: T18267
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1905 <SHA>
A/Cross-references: EMBL:U20432; NID:g664839; PID:g664840; PIDN:AAA62212.1
C:Genetics:
A:Gene: tagB

Query Match 21.9%; Score 492; DB 2; Length 1905;
Best Local Similarity 27.9%; Pred. No. 2.5e-23;
Matches 161; Conservative 73; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYGGQGOIVAVADTGLTGR---NDS-----SMHEAFRGKITALVALGRTNANDTNGH 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 376 LRKGGQILSTADTGLDGHCFSDSKYPIPFQVQVNHKRVTV---YITHDNEYVNGH 432
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 GTHVAGSVLNG-----STNKGMAPQANLVFQSIMDSGGGLGLPSNLQTLFQAYSA 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 GTHVCGSAAGTPEDSSWATSSFSGLATDAKIAFYD-LSSGSSEPTPPEDYSQMYPLXYA 491
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 GARIHNSWCA---AVNGAYTDSRVDDYVRK-NDMTILFAAGNEGNNGTISAPGTA 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 492 GARVHGDSWGSVLOQYGYGSDAGGIDAFIYEPFSLRAAGN-NELFASLAAQATA 550
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 KNAITVGATENLRFSGS-----YADNI----- 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 551 KNAITVGAETAHVNVVSDALEYVDFSDNANFORPCLFDKCYCNVYTTAKCCSEVSNVKG 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 -----NHVAQFSRGPTKDGRIKPDVWAGTILSARSS- 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 611 QLCPPASIKONASDSFTTQPFYNENNGSFSKGPTHDGRKLPDIAVPGVITSARNSG 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 -----LAPDSSFWANHDSKYAVMGTSMATPIVAGNVAQLREHF-----VKNRGI 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 671 ENSTDQCGDSL--PNANGLMSISGTSMATPLATAATTLRQVLVDGYFPTGESVEENKL 728
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 TPPELSLKALAGADIGLGY-----PNGNGWGRVTLDKSLVA 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 729 LP7GSLKALMINNAQLNGTYFWASSTNPSNAIFEQINGANLQGWALRMN---NWL 785
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 YVNESS-----SLTSQKAT-----YSPT-----ATAGK 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 786 YVKSNNPTPSRMWIGGLGKNQKATENKEDSLSSGLNKSICYTYKPPSSSSSGSGGGGT 845
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 P-LKISLWSDAPASTTASVTLVNDLNL-----VITAPN--GTQVGNDFTSVYND 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 846 PRIVATLVMTDPPSYSGAKENLVNLDLNLNDDSDSIITIGNSGSLQFAGKVAQP---- 902
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 NWGGRNVNVEINAPQSGTYTIEVQANVPVGPOTFS 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 903 --DTLNNVEGIIINPTKAMVYKTIAGTNVPIGPQKFS 938
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
B83891
intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (strain
C:Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: B83891
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83891
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-444 <SFO>
A/Cross-references: GB:BA000004; NID:g10174345; PIDN:BA05649.1; GSPDB:GNO
A:Experimental source: strain C-125
C:Genetics:
A:Gene: aprX

Query Match 15.5%; Score 347.5; DB 2; Length 444;
Best Local Similarity 29.3%; Pred. No. 6.1e-15;
Matches 107; Conservative 56; Mismatches 110; Indels 85; Gaps 16;

QY 11 DVAQSSYGLYGGQIVAVADTGLTGRNDSMHEAFRGKITALY-ALGRTNANDTNGHG 69
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 135 EVIENGFTLTKDVTIAVDITGI-----YPHEDLEGRIKAFVDFVNGREEDPDNDNGH 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 THVAGSVLNGSTN---KGMAPQANLVFQSIMDSGGGLGLPSNLQTLFQAYSAAGARI 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 THCGADGAGGASSDGGYRGPAPEANVIGVKVLNK-QMGSLSEIMQGV----- 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 HTNSWGAANVAGYTTDSRVN-----DDYVR-----KNDMTILFAAGN 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 236 ---EMCIQYNEEPDPPDIHISMSLGGQALPYNEQEDPMVRIVEAMNAGITVCVAAGN 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 EGPNGGTISAPGAKNAITVGTATENLRPSGVSADINHVQAPSSRGPTKGRKIPDQVA 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 293 SGPDQAQTIASPGVSEKVIITVGALDD-RDITDREDD--VAPSSRGPTIYGRKPDILA 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 PGTFILSARSLAPDSFS-----WANHDSKYAVMGTSMATPIVAGNVAQLREHFVKNRG 277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 PGYNIVLSRS---PNSFYDKIQGSRVGSHTVMTSGTSMATPVCAGVWALMLQH---EPN 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 ITPKPSLLKALAGADIGLGYPNQNGWGRVTLDKSLNV---AYNNESSSLSTSQK 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 403 LTPDE--VKRLM-----EGDTRWA---DRDPNVYAGYISAFGAIPNSEE 443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
A69587
intracellular alkaline serine proteinase aprX - Bacillus subtilis
C:Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Accession: A69587
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
A.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Etian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.;
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galle
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serc
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yasumoto, K.; Yamoto, K.; Yata, K.; Yoshida,
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69587
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```


R; Voorhorst, W.G.B.; Eagen, R.I.L.; Geerling, A.C.M.; Plattew, C.; Siezen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
A; Title: Isolation and characterization of the hyperthermostable serine protease, pyroly
A; Reference number: Z20481; MUID:196355370; PMID:8702780
A; Accession: T28159
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-1398 <VOO>
A; Cross-references: EMBL:U55835; NID:G1556462; PID:G1556463; PIDN:AA09761.1
A; Experimental source: DSM3638
C; Genes: pls
C; Keywords: hydrolase; serine proteinase

Query Match 13.5%; Score 303.5; DB 2; Length 1398;
Best Local Similarity 26.6%; Pred. No. 1.9e-11;
Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;

Qy 21 GQCIIVAVADTGLDGTGRNDS-----SMHEAFRGKITALYALGRNNAN----- 63
Db 301 GNGYDIAYVDTLDYDFTDEVLQGVNTYDVAVFSYYGPNVYLAIEDPNEGAYAVFGW 360
Qy 64 DTNGHGHVAGSVLNGSTN-----KGMAPQAN 91
Db 361 DGHGHGHVAGVAGVDSNNDWDLMSYSGEWEFSRLYGHYTNVTDTVQGVAPQAQ 420
Qy 92 LVFQISIMDSGGGLGLPSNLQTLFQSAVSAGARIHTNSWGAAGVAYT--TDSRN--VDD 147
Db 421 IMAIRVLR--DGRGSMWDIEGM--TYAATHGADVISMISLGG--NAPYLDGTDPSVAVDE 476
Qy 148 VYRKNDWTLFPAAGNPGNGGTISAPGTAKNAITVAGTENLRPSGSYAD----- 197
Db 477 LTKYGVVFIAGNPGNGINVGSGVATKAITVGAAG--VFVINGVVSQALGVDPDYG 535
Qy 198 -----NINHAQFSSRGPTKGRKPDVNPAGTFFILSARSLAPDSSFWANHDSKYA 249
Db 536 FYVFPAYTNV--RTAFSSRGPRIDGIEKPNVAVGVIGYSSLPWIMGGADF----- 585
Qy 250 YNGGTSMATPIVAGNVAQLREHFVKNRGITPKESLLKALIAAGADI-----GLGYPNG 303
Db 586 -MSGTSMATPHVSGVALLISG--PKPEGIYYPDIKKVLESATWLEGDPYTGOKYTEL 643
Qy 304 NOGWGRVTDKSLNVAVYVNESSLSQKATSFYATAGKPLKISLVNSDAPASTA--- 360
Db 644 DQGHGLVNTKSWEI-----LKAINGTTLPIVDHWADKSYSDFAEYL 685
Qy 361 SVTLVNDLNLVITAPN-----GTQYGN-----DFTSPYNDW-----DG-----RNVENVF 403
Db 686 GVDVIRGLYARNIPDIVEWHIKYVGDTEYRTFEIYATEPWIKPFVSGSVILENNTEFVL 745
Qy 404 -----INAPQSGTY-----TIEVQAVNPVPGPQTF 429
Db 746 RVKYDVEGLEPGLVGRILIDPTTPVIEILNTIVIEKKT 788

RESULT 7
subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)
C; Species: Bacillus sp.
C; Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
C; Accession: S25835
R; Davail, S.; Feller, G.; Narinx, E.; Gerday, C.
Gene 119, 143-144, 1992
A; Title: Sequence of the subtilisin-encoding gene from an antarctic psychrotroph Bacillu
A; Reference number: S25835; MUID:93012966; PMID:1398082
A; Accession: S25835
A; Molecule type: DNA
A; Residues: 1-419 <DAV>
A; Cross-references: EMBL:X63533; NID:G40198; PIDN:CAA45096.1; PID:G40199
C; Superfamily: subtilisin; subtilisin homology
C; Keywords: extracellular protein; hydrolase; serine proteinase
F; 1-23/Domain: signal sequence #status predicted <SIG>
F; 24-110/Domain: propeptide #status predicted <PRO>

F; 111-419/Product: microbial serine proteinase #status predicted <MAT>
F; 135-373/Domain: subtilisin homology <SBT>
F; 144,184,359/Active site: Asp, His, Ser #status predicted

Query Match 12.5%; Score 281; DB 1; Length 419;
Best Local Similarity 33.0%; Pred. No. 1e-10;
Matches 87; Conservative 38; Mismatches 97; Indels 42; Gaps 13;

Qy 21 GQCIIVAVADTGLDGTGRNDSME-EAFRGKITALYALGRT-----FTVGFNFTDSCDTRQGHGHVAGSV 76
Db 135 GAGINTAVLDTGVNTNHPDLSNNVEQCK-----

Qy 77 LKNGSTNK-----GMAPOANL-VFOSIMDSGGGLG-GLPSNLQTLFQSAVSAGARIHTN-SW 130
Db 190 LANGGTGSGVYGAPEADLMAYKVLGDDSGYADDAEAIHAGDOATALNTKVINMNL 249

Qy 131 GAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAITVAGTENLRP 190
Db 250 GSGGESLIT---NAVYAYDKGVLLIAAGNSGPKPGSIGYPGALVNAVVAALNTTIQ 306

Qy 191 SFGSYADNINHVAQFSRSGPTKDG-----RIKPDVNPAGTFFILSARSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSRGHKRTAGDYVIQGDVEISAPGAAYST-----W--F 348

Qy 245 DSKYAYMGTSMATPIVAGNVAQL 268
Db 349 DGGYATISGTSMAAPRAAGLA 372

RESULT 8

G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodura
C; Species: Bacillus halodurans
C; Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C; Accession: G83753
R; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A; Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A; Reference number: A83650; MUID:20512582; PMID:11058132
A; Accession: G83753
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-799 <STO>
A; Cross-references: GB:AP001510; GB:BA000004; NID:G10173440; PIDN:BA04550.1; GSPDB:GNO
A; Experimental source: strain C-125
C; Genes: vpr
A; Superfamily: microbial serine proteinase vpr; subtilisin homology
C; Keywords: hydrolase; serine proteinase
F; 1-29/Domain: signal sequence #status predicted <SIG>

Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 2.8e-10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;

Qy 18 GLYQGGQIVAVADTGLDGTGRNDSMEAFRGKITALYALGRNNANDT-----NGH 68
Db 171 GYTGEGITVAITDGTVDYTHFD--LVHAF-GDYKGMDFIDNDDPQETPPGDPGRGIETH 227

Qy 69 GTHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGLPSNLQTLFQSAVSAGARIHTN 128
Db 228 GTHVAGVVAAGLI-KGVAPDANLLAYRLVLPQGG--RGSTAGVIAGIERAVQGDIMNL 284

Qy 129 SWGAANVAVYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAITVAGTENL 188
Db 285 SLGNTLNDPDPFATSIAL-DWMAEAGVAVTNSGNSPNWTVSGSPGTSRDAISVGAT--- 340

Qy 189 RPSFGSY-----ADNINH----- 201
Db 341 RLPNYTKASVFTSDGIDYPSADIMGPPSDEELLELDGETYEFYAFGLKPGDFEGVDVE 400

Qy 202 ----- 201

Db 401 GKIALIVRGSIPTVEKAENAKAGVGAIIYNNVAGVQVTPGLAIPITIMLSNEDGLKVR 460
QY 202 -----VAQSSRGPT-KDGRIKPDVMAFGTIFLSARSGLAPD 237
Db 461 NELENGQNTVTFSEBPKLVGETVADFSGRPVMTWMIKPDVSPGVAIVSTIPHQPD 520
QY 238 SSWANHDSKIAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAAALIAGAADI- 296
Db 521 DPT-----GYGSRQGTSMASPHVGAALLLEAH-PNWGV---DHVKAALMTAENLV 569
QY 297 ---GLYFPGNQGWRVTLDKSLNVAYNVSSLSLSQKATY-SPTATAGKPLK 346
Db 570 DENGNYPHNTQAG-----SIRIVDAIESETLVTPGSHSGFTFKERKQVE 617

RESULT 9
T29090
surface layer-associated STABLE proteinase - Staphylothermus marinus
N/Alternate names: hyperthermostable proteinase
C/Species: Staphylothermus marinus
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T29090
R/Mayr, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A/Title: A hyperthermostable protease of the subtilisin family bound to the surface layer
A/Reference number: Z20559; MUID:96385442; PMID:8793300
A/Accession: T29090
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1345
A/Cross-references: EMBL:U57968; NID:G1374755; PID:G1374756; PIDN:AAB02323.1
A/Experimental source: strain F1
C/Function:
A/Description: probably serves an exodigestive function related to the organism's energy
A/Note: stoichiometric S-layer component

Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.2e-10;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;
QY 46 FRKITALYALGRNNDNTHNGHTVA-----GSVL-----QNGSTNK--GNAPQANLV 93
Db 445 YQGRYAL-----VSDPHGHGTSVATVIAASGRVLYDYGDKLYRIMGVAPGAKI- 495
QY 94 FQSIMDSGGGLGIPSNLQTLFQVAGSAG-----ARHTNSM 130
Db 496 -----AGDALLGNILV--EAWLAGNIVTEEDGVVYLSLDPFGPHRADIIISNW 546
QY 131 GAAVNGAYTTDSRNV-----DYVRKNDMTILFAAGNEGPNNGTISAPGTAX 177
Db 547 GSIYINFLQFPDIDYRSSFMDDEILAIRNYLIGDHVTIVFAAGNEGPGYSSNGAPGTGL 606
QY 178 NAITVGATE--NLRPSFG---SYADNINHVAQFSSRGPTKDGRIKPDVMAQTFILSARS 232
Db 607 LVITAGASTLWDYTRIYGYEGVAD---EVIPSSRGPTQGGYKPKDPIVNIAGFEWASTR 663
QY 233 SLAPDSSFWANHDSKIAYMGTSMTATPIVAGNVAQLREHFVKNGRITPKPSLLKAAALIAG 292
Db 664 TI-DGRGYGAQPD---VFGGTSEATPYTSGTLALVQAYKEVYNTTPDPVTAKILKSS 718
QY 293 AADIGLYPNGNQGWRVTLDKSLNVAYNVE 323
Db 719 AKDI--WYPAFSGSGRVDALKAADTVFISE 747

RESULT 10
S11890
serine proteinase (EC 3.4.21.-) precursor, extracellular - Xanthomonas campestris pv. ca
N/Alternate names: subtilisin-related proteinase
C/Species: Xanthomonas campestris pv. campestris
C/Date: 21-Nov-1993 #sequence_revision 07-Feb-1997 #text_change 03-Dec-1999
C/Accession: S11890
R/Liu, Y.N.; Tang, J.L.; Clarke, B.R.; Dow, J.M.; Daniels, M.J.

Mol. Gen. Genet. 220, 433-440, 1990
A/Title: A multipurpose broad host range cloning vector and its use to characterise an e
A/Reference number: S11890; MUID:90251253; PMID:2187155
A/Accession: S11890
A/Molecule type: DNA
A/Residues: 1-580 <LIU>
A/Cross-references: EMBL:X51635; NID:948533; PIDN:CAA35962.1; PID:948534
A/Experimental source: Xanthomonas campestris pv. campestris
A/Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 205-Ala
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F/1-32/Domain: signal sequence #status predicted <Sig>
F/168-423/Domain: subtilisin homology <SET>

Query Match 12.4%; Score 278; DB 2; Length 580;
Best Local Similarity 26.0%; Pred. No. 2.5e-10;
Matches 126; Conservative 56; Mismatches 147; Indels 156; Gaps 24;
QY 21 GQGIIVAVDTGL-----DTGRNDSMHEAFRGKITALYALGRNNDN----- 64
Db 168 GSGTVAVIDTGITSHADLANILAGYDFISDATTARDGNGRDSNAADGDMVAANECCA 227
QY 65 -----TNGHGTTHVAGSVLGNSTNGKMAPOA-----NLVFQSIMD 99
Db 228 GIPAASSWHGTHVAGIVAAVTNTTGVAGTAYGAKVVPVVLGKCGSLSDIADAIWA 287
QY 100 SGGGLGGLPSNLQ--TLFSQAYSAGARIHNSWGAANGAYTTDSRNVDYVRKNDMTIL 157
Db 288 SGGTVSGIPANANPAEIVNMSLGGGSCSTTMQN-AINGAVSRGT-----TVV 334
QY 158 FAAGNEGPNNGTISAPGTAKNAITVGATEN--LRPFSGSYADNINHVAQFSSRGPTKDG 215
Db 335 VAANDASNVSG-SLPANCANVIAVATTSAGAKASINFGTGI----- 377
QY 216 IKPDVMAQTFILSARS--LAPDSSFWANHDSKIAYMGTSMTATPIVAGNVAQLRHFV 273
Db 378 ---DVSAPGSSILSTLNSGITTPGS-----ASYASNGTSMASPHVAGVVALVQS--V 425
QY 274 KNRGITPK--PSLLK--AALIAGRADIGLYPNGNQGWRVTLDKSLNVAYNVSS----- 325
Db 426 APTALTPAAVETLLKNTARALFGAC-----SGCGAGIVNADAAYTAA-INGSGGGG 477
QY 326 -----SLSTQKATYSTATAGKPLKLSLVMSDAPASTTASVTL---VND 367
Db 478 GGGNTLTNGTPTVGLGAATGAELNYTITVPAG-----SGTLTVTSSGSGD 523
QY 368 LNLVI---TAPNGTQYVGNDFTSYNDNWDGRNVENVFINAPQSGTYTTEVQAYNYPVG 424
Db 524 ADLYVRAGSAPTDSAYT-----CRPYRS-----GNAETCTITAP-SGTYYVRLKAYS----- 569
QY 425 PQTFES 429
Db 570 --TFES 572

RESULT 11
S23407
subtilisin (EC 3.4.21.62) 1 precursor - Bacillus sp. (strain TA39)
C/Species: Bacillus sp.
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C/Accession: S23407
R/Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A/Title: Nucleotide and derived amino acid sequence of the subtilisin from the antarctic
A/Reference number: S23407; MUID:92256481; PMID:1581352
A/Accession: S23407
A/Molecule type: DNA
A/Residues: 1-420 <NAP>
A/Cross-references: EMBL:X62369; NID:940200; PIDN:CAA44227.1; PID:940201
C/Genetics:
A/Gene: sub1
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase

F;1-23/Domain: signal sequence #status predicted <SIG>
 F;24-111/Domain: propeptide #status predicted <PRO>
 F;112-420/Product: microbial serine proteinase #status predicted <MAT>
 F;136-374/Domain: subtilisin homology <SBT>
 F;145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
 Best Local Similarity 31.2%; Pred. No. 1.9e-10;
 Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GGGQIVAVADTGLDGRNDSVHEAFRGKITAL--YALGRT---NNANDTGHGTHVAGS 75
 DB 136 GGGINIAVLDTGNTN-----HPDLNNVECKFTVGTITNNSCCTDRQGHGTHVAGS 189
 QY 76 VLNGSTNK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLCTLPFSAGARIHTN-S 129
 DB 190 ALADGGTGNVGVAPADADLWAYKVLGDDSGYADDIAAIRHAGDQATALNTKVINNS 249
 QY 130 WGAANVGATTSRNVDDYVRKNDMTILPAAGNEGNGTISAPGTAKNAITVGATENLR 189
 DB 250 LGSSESSITNAVN---YSYKNGVLITAAAGNSGPGYQSGIYPGALVNAVAVALEN-K 305
 QY 190 PPSGYSADNINHVAQFSSRGPT-KDG-----RIKPDVMAPGTFILSARSLAPDSFWAN 243
 DB 306 VENGTY-----RVADFSSRGYSWTGDYAIQKGDVEISAPGAIIYST-----N-- 348
 QY 244 HDSKIAYMGSTWATPIVAGNVAQLREHFVKRGITPKPSLLKALIAAGADIGLGPNG 303
 DB 349 FDGGYATISGTSNASPHAAGLAAKTIWAQYSPASNVDRVGELOQYRAY---ENDILSGYYAG 405
 QY 304 -----NQGWGRVTL 312
 DB 406 YGDDFASGFGFATV 419

RESULT 12
 JC4908
 alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
 N;Alternate names: subtilase
 C;Species: Alteromonas sp.
 C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
 C;Accession: JC4908
 R;tsujiibo, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
 Biosci. Biotechnol. Biochem. 60, 1284-1288, 1996
 A;Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas*
 A;Reference number: JC4908; MUID:97141200; PMID:8987544
 A;Accession: JC4908
 A;Status: Preliminary
 A;Molecule type: DNA
 A;Residues: 1-715 <TSU>
 A;Cross-references: DDBJ:D38600; NID:gl536787; PIDN:BAAL8912.1; PID:d1019647; PID:g21602
 A;Experimental source: strain O-7
 C;Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensitized subtilisin.
 C;Genetics:
 A;Gene: aprI
 C;Superfamily: subtilisin homology
 C;Keywords: hydrolase
 F;1-40/Domain: signal sequence #status predicted <SIG>
 F;41-150/Domain: amino-terminal propeptide #status predicted <ATP>
 F;151-496/Product: alkaline serine protease
 F;182-452/Domain: subtilisin homology <SBT>
 F;497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F;739-294,335-372,478-481/Disulfide bonds: #status predicted

Query Match 11.8%; Score 265; DB 2; Length 715;
 Best Local Similarity 25.8%; Pred. No. 2.2e-09;
 Matches 124; Conservative 48; Mismatches 170; Indels 138; Gaps 22;

QY 21 GGGQIVAVADTGLDGRNDSVHEAFRGKITALYALGRNAND-----TNG---- 57
 DB 182 GGGVAVAVLDTGYPHLLDANILPGYDMISNTFVANDGGARDNDARDPGDAVTRECGT 241
 QY 68 -----HGTHVAG---SVLNGSGTNKGMAPQANLVFQSIMDSGGGLGLP 108

Db 242 DSSGQPVPRADQDSSWHGTHVAGTAAATNNGEVAGVAYDAKVVPRVL---GKCGGLT 298
 QY 109 SNLQTLFSQAYSAGARIHTNSWGAAY-----NGAYTTDSRNVDDYVRKNDMTILFAAG 161
 Db 299 SDIADGLIIVASGGSDRVPANPVPVNNMSLGGGACSAATTQVAINQARNNGTVIVIAG 358
 QY 162 NEGPNGGTISAPGTAKNAITVGAT--ENLRPFGSGYADNINHVAQFSSRGPTKGRIPKD 219
 Db 359 NDNDNSANTY-PCNCNGVNVASVGRDGRAYVSNYGANI-----D 398
 QY 220 VMAPGTFILSARSLAPDSFWANHDS-----KYAYMGSTMATPIVAGNVAQLR-- 269
 Db 399 VAAPG---GAQSFADDEGILSTHNSGSGAPENDSYHSQGTSMAPHPVAGVAAALIKQA 454
 QY 270 -----EHFYKN--RGITPKPSLLKALIAAGADI--GLG-----YPNGMGWGRVTL 312
 Db 455 KPSATPDEVETILKNTTRSPAGSCSCGTGTVWDAANVAALGDVVPPTGN-----TL 508
 QY 313 DKSLNVAYNNESSLSSTSQKATYSFTATAGKPKLSLWMSDAPASTASVTLVNDLNLVI 372
 Db 509 ED--GVAKTGLSGAAGSNQFFTF-----NNVENVFINAPQSGTYTIEVQAYNPVG 540
 QY 373 TAPNGT---QYVGNDFTSYNDNDWGR---NNVENVFINAPQSGTYTIEVQAYNPVG 424
 Db 541 TMSGTGADLYVKLG-SQPTSSSYDCRPVEGGNAEVCSPADAPQAGTYHWMINGYKAYSG 599

RESULT 13
 A72647
 Probable surface layer-associated STABLE proteinase APE0607 - Aeropyrum pernix (strain A72647)
 C;Species: Aeropyrum pernix
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C;Accession: A72647
 R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka-awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum*
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: A72647
 A;Status: Preliminary
 A;Molecule type: DNA
 A;Residues: 1-1331 <RAW>
 A;Cross-references: DDBJ:AF000060; NID:95104186; PIDN:BAA79577.1; PID:d1043363; PID:g51
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE0607

Query Match 11.7%; Score 264; DB 2; Length 1331;
 Best Local Similarity 23.6%; Pred. No. 6e-09;
 Matches 111; Conservative 66; Mismatches 158; Indels 136; Gaps 18;

QY 8 VKADVAOSSGLYQGGQIVAVADTGLDGRNDSVHEAFRGKITALYA----- 55
 Db 346 VYADLS-TAYYLF-----LKALSDTGMTSGEPDPSLLDLSFADETPASYGSEVLARDPTGD 400
 QY 56 -----LGRTNAN-----DTNGH 68
 Db 401 GVNDFSAGALAGTYDWVGLLTGSEVNLGWRGLGFDYAGLVPLGLDPQGRWVSLYDTLAH 450
 QY 69 GTHVAGSVLNGSTN-----KGMAPQANLVFQSIMDSGGGLGLPSNLQTLFSQA 118
 Db 461 GTSVATVIAIRGNVFNGLGYIETSLRGVAFGAKIA-----AGGSF-----LINVFAQL 509
 QY 119 YSAG-----ARIHTNSW---AAVNGAYT---TDSRNVDDY-VRKNDMT 155
 Db 510 FLSGFEPQDSPNLNVVYTGHEQVDVNNWNSYIALRGFLTGADYATIEDYIVSAGTV 569
 QY 156 ILFAAGNEGNGGTISAPGTAKNAITVGATE--NLRFSGYSADNINHVAQFSSRGPTKD 213
 Db 570 IVHANGNGGFOYGTATTPGAGSLIISVGASTLDFRYFYGLPSPGCDVLSWDRGFSQI 629
 QY 214 GRIKPDYMAPGTFILSARSLAPDSFWANHDSKIAYMGSTMATPIVAGNVAQLREHFV 273

Db 630 GVAKPDVNTGSAWAG-----VPVLTGLGNSLAFDIFGCTSEATPTSGSVALVISAYQ 685
Qy 274 KYRGIPEKPSILKAAIAGAADIGLGYPNGNQGRVTLDDKSLNV-----AY 320
Db 686 QAFGAKESPGLVKAILAKSTARDTGA--DAFTQSGGVDDVYRAVKAVLEGVPIALSTSVY 743
Qy 321 VNSSSLSTSQKATYSTATAGKPLKLSLWSDA--PASTASVTLVNDLN 369
Db 744 ENVYSLLS-----GYSYPPFLAPNPVEDTQYPGVLKPEGETAVETLVKLTL 789
RESULT 14
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C/Species: Thermus aquaticus
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: A35742; S00620; S00324
R/Takami, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6591, 1990
A/Title: Unique precursor structure of an extracellular protease, aqualysin I, with NH-2
A/Reference number: A35742; MUID:90216674; PMID:2182621
A/Accession: A35742
A/Molecule type: DNA
A/Residues: 1-513 <TER>
A/Cross-references: GB:J90108; GB:D90108; NID:g217171; PIDN:BAAL4135.1; PID:Q
A/Note: the authors translated the codon CTG for residue 470 as Val, and GGT for residue
R/Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 431-437, 1988
A/Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine
A/Reference number: S00620; MUID:98225062; PMID:3286255
A/Accession: S00620
A/Molecule type: DNA
A/Residues: 75-442 <KWO>
A/Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A/Note: part of this sequence, including the amino and carboxyl ends of the mature prote
R/Matsuzawa, H.; Tokugawa, K.; Hamakoshi, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; Kwon
Eur. J. Biochem. 171, 441-447, 1988
A/Title: Purification and characterization of aqualysin I (a thermophilic alkaline serin
A/Reference number: S00324; MUID:98151937; PMID:3162211
A/Accession: S00324
A/Molecule type: Protein
A/Residues: 128-170 <MATS>
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F/1-14/Domain: signal sequence #status predicted <SIG>
F/15-127/Domain: propeptide #status predicted <PRO>
F/128-408/Product: aqualysin I #status experimental <MAT>
F/157-364/Domain: subtilisin homology <SBT>
F/255-257,281-283/Region: SI specificity crevice #status predicted
F/409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F/166,197,349/Active site: Asp, His, Ser #status predicted
Query Match 11.7%; Score 262.5; DB 1; Length 513;
Best Local Similarity 26.2%; Pred. No. 2.1e-09;
Matches 118; Conservative 43; Mismatches 144; Indels 145; Gaps 23;
Qy 16 SYGLYGGQIVAVADTGLDTRNDSSMHEAFRKITALLY-ALGRNNANDTNGHGHVAG 74
Db 152 TYTATGRGVNVYVDTGIRT-----THREFGGRARVGYDALG--GNGQDCNGHGHVAG 203
Qy 75 SVLNGSTNGKMAPOANLVFQSIMD--SGGLGLPSNLQTLFSAQYAGARHTN- --- 128
Db 204 TI---GGVTYGVAVKAVNLVAVRLDNCNGSGSTSGVIAGVDWV-----TRHRRPAVA 252
Qy 129 --SWGAAVNGAYTTDSRNVDYVRKN---DMTILFAAGNEGPNGGTISAFGTAKNAITVG 183
Db 253 NMSLGGGVSTA-----LDNAVKNISIAAGVYVAAGNDNACNYS-PARVAEALTVG 304
Qy 184 AT--ENLRPSFGSYADNINHVAFSSRGPTKGRIPDVMAPGCTFILSARSSLAPSSFW 241
Db 305 ATTSARASFSNYGSCV-----DLFAPGASIPSA-----W 335

Qy 242 ANHDSKYAMGCGTSMATPIVAGNVAQLREHFVKNRGITP-----KPSLLKAAIAGAADIGL 298
Db 336 YTSDTATQTLNGTSMATPHVAG-VAAL--YLEQNPSATPFASVAILNGATTGRLSGIGS 392
Qy 299 GYPNGNQGRVTLDDKSLNVAVNNESSLSLSQKATYSTATAGKPLKLSLWSDAPAST 358
Db 393 GSPN-----RLLYSLSSGS-----GSTAPCTS 415
Qy 359 ---TASVTLVNDNLVITAPNGTQY-----VGNDFTSYNDNWDGRN--- 397
Db 416 CSYTTGSLSGPDYNF---QPNGTVYYSYPAGTHRAWLRGPAGTDF-DLYLWRWDGSRWLT 471
Qy 398 -----NVENVFINAPQSGTYTIEVOAYN 420
Db 472 VGSSTGPTSEESLSYSGTAGYLLWRIYAYS 501
RESULT 15
C84120
subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodi
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: C84120
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: C84120
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-757 <STO>
A/Cross-references: GB:AP001519; GB:BA000004; NID:gi0176109; PIDN:BA07482.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH3763
C/Superfamily: microbial serine proteinase vpr; subtilisin homology
C/Keywords: hydrolase; serine proteinase
F/1-20/Domain: signal sequence #status predicted <SIG>
Query Match 11.4%; Score 256.5; DB 2; Length 757;
Best Local Similarity 22.2%; Pred. No. 8.5e-09;
Matches 117; Conservative 59; Mismatches 157; Indels 193; Gaps 15;
Qy 8 KVADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRG-----KITALLYALG 57
Db 119 VEGMLDEEGVHLTGKGVKVAVIDTGDYTHPD--LQSSYKGYDFVDYDDDPMETIASQG 176
Qy 58 RTNNANDTNGHGTAVAGSVLNGSTNGKMAPOANLVFQSIMDSSGGLGLPSNLQTLFSQ 117
Db 177 -----PTLHGTHVSGIIAANGQV-KGVAPAEIYAYRALCPGG--QGTEQVIAAEK 227
Qy 118 AYSAGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAK 177
Db 228 AVEDGVDVNLSLGNTVNGPDWPTSLADAAVEGVAVT-SNGNSGPNMWTGSPGTSK 286
Qy 178 NAITVGAT-----EN--LRPSFGS----- 194
Db 287 KAISVGASAPPLNTYLTAFGEENEISLYPSSGGLPWAFKDLPMIDVGYGTEKEWGVGD 346
Qy 195 -----YADNINH----- 201
Db 347 AEGKVVLKRGVWPFTEKVMHAAKARGVLIYNNTPGPTGMEGVNIPVWSITREDG 406
Qy 202 -----VAQSSRGP-TKGRIKPDVWAPGCTFILSARSS 233
Db 407 BFLLEQLQKNKELTLTIYRKEEDFVALFESSRGVPTHTWTDVKPDVWVAVGSI- ---DS 462
Qy 234 LAPSSFWANHDSKYAVMGGTSMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAAI 291
Db 463 TIPNG-----YLGNGTSMAPHVAGAAALIKQ-----AHEWTEPEQVKAALMN 507
Qy 292 GAADI-----GLGYNGNQGRVTLDDKSLNVAVNNESSLSLSQKATYSTATAGKPLKI 347

Db 508 TAKKLVQEGVPHEIHEQGAGRIQVDKAV-----AATSLVYFGALSFGK---- 551
QY 348 SLVMSDAPASTTASVTLVNDLNLVITAPNGTOYVGNFTSPYNDNW 393
Db 552 ---WSKDDLREKEPVTLTIENTHDTV--KRTYHISPPFDVDPDGVEM 591

Search completed: March 10, 2004, 14:55:11
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:47:28 ; Search time 17 Seconds
(without alignments)
1329.320 Million cell updates/sec

Title: US-09-985-689A-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EYQAVNPVGPQTFLAIWN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	519.5	23.1	1743	1	TAGC_DICDI
2	492	21.9	1905	1	TAGB_DICDI
3	311.5	13.9	806	1	SUBV_BACSU
4	305.5	13.6	1398	1	PLS_PYRFU
5	278	12.4	580	1	EXPR_XANCP
6	277	12.3	420	1	SUBT_BACS9
7	262.5	11.7	513	1	AQLI_THEAQ
8	253.5	11.3	894	1	WPRB_BACSU
9	246.5	11.0	402	1	ALP_CEPAC
10	244.5	10.9	534	1	PROA_VIBAL
11	242.5	10.8	401	1	THES_BACSP
12	240	10.7	269	1	SUBS_BACLE
13	240	10.7	380	1	ELYA_BACAO
14	240	10.7	380	1	ELYA_BACCS
15	239	10.6	269	1	PRTM_BACSP
16	235.5	10.5	382	1	SUBT_BACAM
17	235	10.5	378	1	ELYA_BACSP
18	232	10.3	269	1	SUBB_BACLE
19	231	10.3	321	1	ISP_BACCS
20	230.5	10.3	379	1	SUBT_BACLI
21	230	10.2	404	1	SMPI_MAGFO
22	227	10.1	1181	1	SCA2_STRPY
23	226.5	10.1	1052	1	MS1P_CRIGR
24	226.5	10.1	1052	1	MS1P_HUMAN
25	226.5	10.1	1052	1	MS1P_MOUSE
26	226.5	10.1	1052	1	MS1P_RAT
27	225.5	10.0	381	1	SUBN_BACNA
28	225.5	10.0	1167	1	SCAL_STRPY
29	223.5	9.9	381	1	SUBT_BACSA
30	223.5	9.9	381	1	SUBT_BACST
31	223.5	9.9	381	1	SUBT_BACSU
32	221.5	9.9	275	1	SUBT_BACPU
33	216.5	9.6	274	1	SUBD_BACLI

34	216.5	9.6	530	1	HLV_HALI7
35	216.5	9.6	645	1	SUBB_BACSU
36	215	9.6	422	1	TKSU_PYRKO
37	214.5	9.5	361	1	ELYA_BACHD
38	214	9.5	279	1	THEY_THEVU
39	214	9.5	293	1	PRTT_TRIAL
40	214	9.5	326	1	ISP_FAEPO
41	214	9.5	409	1	ALP_TRIHA
42	211	9.4	319	1	ISPI_BACSU
43	210.5	9.4	1433	1	SUBF_BACSU
44	208.5	9.3	388	1	CUDP_METAN
45	208	9.3	387	1	PRTR_TRIAL

ALIGNMENTS

RESULT 1

TAGC_DICDI	AC	Q23868	DT	01-NOV-1997	(Rel. 35, Created)
TAGC_DICDI	AC	Q23868	DT	01-NOV-1997	(Rel. 35, Last sequence update)
TAGC_DICDI	AC	Q23868	DT	15-MAR-2004	(Rel. 43, Last annotation update)
TAGC_DICDI	AC	Q23868	DE	Prestalk-specific protein tagC precursor (EC 3.4.21.-).	
TAGC_DICDI	AC	Q23868	GN	TAGC.	
TAGC_DICDI	AC	Q23868	OS	Dictyostelium discoideum (slime mold).	
TAGC_DICDI	AC	Q23868	OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.	
TAGC_DICDI	AC	Q23868	OX	NCBI_TaxID=44689;	
TAGC_DICDI	AC	Q23868	RN	[1]	
TAGC_DICDI	AC	Q23868	RP	SEQUENCE FROM N.A.	
TAGC_DICDI	AC	Q23868	RC	STRAIN=AX4;	
TAGC_DICDI	AC	Q23868	RX	MEDLINE=97140317; PubMed=986798;	
TAGC_DICDI	AC	Q23868	RA	Shauly G., Escalante R., Loomis W.F.;	
TAGC_DICDI	AC	Q23868	RT	"Developmental signal transduction pathways uncovered by genetic suppressors";	
TAGC_DICDI	AC	Q23868	RL	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).	
TAGC_DICDI	AC	Q23868	CC	-!- FUNCTION: Intercellular communication via tagC may mediate integration of cellular differentiation with morphogenesis (By similarity).	
TAGC_DICDI	AC	Q23868	CC	-!- SIMILARITY: In the N-terminal section; belongs to peptidase family S8.	
TAGC_DICDI	AC	Q23868	CC	-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.	
TAGC_DICDI	AC	Q23868	CC	-!- SIMILARITY: STRONG, TO TAGC.	
TAGC_DICDI	AC	Q23868	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
TAGC_DICDI	AC	Q23868	CC	EMBL; U60086; AAB03331.1; --	
TAGC_DICDI	AC	Q23868	DR	PIR; T18279; T18279.	
TAGC_DICDI	AC	Q23868	DR	DictyBase; DDB0001795; tagC.	
TAGC_DICDI	AC	Q23868	DR	InterPro; IPR003593; AAA_Atpase.	
TAGC_DICDI	AC	Q23868	DR	InterPro; IPR001140; ABC_TM_transpt.	
TAGC_DICDI	AC	Q23868	DR	InterPro; IPR003439; ABC_transporter.	
TAGC_DICDI	AC	Q23868	DR	InterPro; IPR002099; Peptidase_S8.	
TAGC_DICDI	AC	Q23868	DR	Pfam; PF00664; ABC_membrane; 1.	
TAGC_DICDI	AC	Q23868	DR	Pfam; PF00005; ABC_tran; 1.	
TAGC_DICDI	AC	Q23868	DR	Pfam; PF00082; Peptidase_S8; 1.	
TAGC_DICDI	AC	Q23868	DR	PRINTS; PR00723; SUBTILISIN.	
TAGC_DICDI	AC	Q23868	DR	SMART; SM00382; AAA_1	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00929; ABC_TM1F; 1.	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00137; SUBTILASE_HIS; 1.	
TAGC_DICDI	AC	Q23868	DR	PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.	
TAGC_DICDI	AC	Q23868	DR	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;	

KW SIGNAL. 1 27
 FT CHAIN 28 1743
 FT DOMAIN 346 642
 FT TRANSMEM 1450 1687
 FT TRANSMEM 1027 1982
 FT TRANSMEM 1072 1047
 FT TRANSMEM 1157 1092
 FT TRANSMEM 1157 1177
 FT TRANSMEM 1250 1280
 FT TRANSMEM 1288 1308
 FT ACT SITE 325 325
 FT ACT SITE 372 372
 FT ACT SITE 637 637
 FT NP BIND 1485 1492
 FT DOMAIN 42 46
 FT DOMAIN 94 103
 FT DOMAIN 643 646
 FT DOMAIN 733 741
 FT DOMAIN 786 792
 FT DOMAIN 1337 1340
 FT DOMAIN 1346 1352
 FT DOMAIN 1353 1357
 FT DOMAIN 1358 1364
 FT DOMAIN 1381 1386
 FT DOMAIN 1707 1729
 FT CARBOHYD 390 390
 FT CARBOHYD 536 536
 FT CARBOHYD 547 547
 FT CARBOHYD 614 614
 FT CARBOHYD 689 689
 FT CARBOHYD 735 735
 FT CARBOHYD 741 741
 FT CARBOHYD 832 832
 FT CARBOHYD 887 887
 FT CARBOHYD 1251 1251
 FT CARBOHYD 1385 1385
 FT CARBOHYD 1386 1386
 FT CARBOHYD 1454 1454
 FT CARBOHYD 1704 1704
 SQ SEQUENCE 1743 AA; 12DB3362F729839 CRC64;

Query Match 23.1%; Score 519.5; DB 1; Length 1743;
 Best Local Similarity 28.4%; Pred. No. 3.8e-26;
 Matches 166; Conservative 78; Mismatches 158; Indels 183; Gaps 24;

QY 19 LYGGQIVAVATGLDTRG---NDS-----SMHEAFRGKITALYALGRNNANDTNKH 68
 DB 314 LRKGQILSIADTGLDGSCHFFSFKYPIPLNSVNLNR-KVTVTTTSTSDSDSKVDGH 372
 QY 69 GTHVAGSVLG-----NGSTKMGAPQANLVFQSIMDSGGGLGL--PSNLQTLFQAY 119
 DB 373 GTHICGSAAGTEDSSVNISSFGSLATDAKIAF---FDLASGSSSLTFPSDLKQLYQPLY 429
 QY 120 SAGARIHTNSGA---AVNGAYTTDSRNVDDYVRKN-DMTILFAAGNEGNGGTIS--A 172
 DB 430 DAGARVHCDSWGSVSVEGYTGSVSDTASIDDFLETHPFIILRAAGN---NEQYLSLLT 486
 QY 173 PGTAKNAITVGATENLR-----PSFGSYADNI----- 199
 DB 487 QSTAKNVIIVGAHQTHENYLTGDPNYINIQSVSDINQELICDFSRVCNYTTAOCCLLES 546
 QY 200 -----NHVAQFSRSGFTKGRIPKPDVMAFGTFIL 228
 DB 547 NATTGLASCCPTLLRKSVIDAANTQPLLYNENNICSFSSKGFTHGRMKPALVAPGEVIT 606
 QY 229 SARSSIA-----PDSSFWANDSKYVNGGTSMATPIVAGNVAQIREH-----F 272
 DB 607 SAKSNGANTDQCGDGLS-PNTNALLA--ISGTSMATSFATAATTILRQYLVDDGYPTGSI 664
 QY 273 VKNRGRTPKPSLLKAALIAGA-----ADIGLGYPNGN-----QGWGRVT 311
 KW Signal.

Db 665 VESNKLQPTGSLKALMINNAQLNGTFLQITSSITYPSNQVFNFAAGSLVQGWGAIR 724
 QY 312 LDXSINVAYNES-----SLSTSOKATYSFT-- 338
 Db 725 MSNWLHVNNNNNNNNKTSKGITKPDGGDLRLVKPNQWKEBSLSTGQNTSYFTYK 784
 QY 339 -----ATAGKPLK---ISLWSDAPASTTASVTLVN--DLNLVITAFNG-TQYVGNDFTS 387
 Db 785 PSSSSSSNGNNIPRVVATLVWTDPPSYAGAKFNLVNLDLTWIIYRDNGSTIFYSNQGS 844
 QY 388 PY---NDNWDGRNVNENFINAPQSGTYTIEVQAVNVPVGPQTF 429
 Db 845 SFLGLAPTQDTLNNVEGIVHNPTPTMPTFMVAGTNVPMGPONFS 889
 RESULT 2
 TAGB DICDI
 ID TAGB DICDI STANDARD; PRT; 1905 AA.
 AC P34683;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Prestalk-specific protein tagB precursor (EC 3.4.21.-).
 GN TAGB
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=AX4;
 RL MEDLINE=95262903; PubMed=7744252;
 RA Shaulsky G., Kuspa A., Loomis W.F.;
 RT "A multidrug resistance transporter/serine protease gene is required
 for prestalk specialization in Dictyostelium.";
 RL Genes Dev. 9:1111-1122(1995).
 CC -|- FUNCTION: Intercellular communication via tagB may mediate
 integration of cellular differentiation with morphogenesis.
 CC -|- SIMILARITY: In the N-terminal section; belongs to peptidase family
 S8.
 CC -|- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
 TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC -|- SIMILARITY: STRONG, TO TAGC.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DB EMBL; U20432; AAA62212.1; -.
 DB PIR; T18267; T18267.
 DB MEROPS; S08.UPW;
 DB DictyBase; DB0001964; tagB.
 DB InterPro; IPR003593; AAA_ATPase.
 DB InterPro; IPR001140; ABC_TM_transport.
 DB InterPro; IPR003439; ABC_transporter.
 DB InterPro; IPR000209; Peptidase_S8.
 DB Pfam; PF00664; ABC_membrane_1.
 DB Pfam; PF00005; ABC_tran; 1.
 DB Pfam; PF00082; Peptidase_S8; 1.
 DB PRINTS; PR00723; SUBTILISIN.
 DB ProDom; PD000006; ABC_transporter; 1.
 DB SMART; SM00382; AAA; 1.
 DB PROSITE; PS00929; ABC_TM1F; 1.
 DB PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DB PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DB PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
 DB PROSITE; PS00137; SUBTILASE_HIS; 1.
 DB PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
 Signal.

FT SIGNAL 1 31
FT CHAIN 32 1905
FT DOMAIN 378 700
FT TRANSFER 1518 1756
FT TRANSFER 1011 1031
FT TRANSFER 1076 1096
FT TRANSFER 1121 1141
FT TRANSFER 1210 1230
FT TRANSFER 1309 1329
FT TRANSFER 1332 1352
FT ACT_SITE 387 387
FT ACT_SITE 432 432
FT ACT_SITE 695 695
FT NP_BIND 1553 1560
FT DOMAIN 63 67
FT DOMAIN 95 104
FT DOMAIN 107 134
FT DOMAIN 311 321
FT DOMAIN 833 837
FT DOMAIN 838 844
FT DOMAIN 871 876
FT DOMAIN 1012 1015
FT DOMAIN 1386 1389
FT DOMAIN 1398 1404
FT DOMAIN 1445 1450
FT DOMAIN 1765 1779
FT DOMAIN 1782 1785
FT DOMAIN 1807 1812
FT DOMAIN 1813 1860
FT DOMAIN 1872 1878
FT CARBOHYD 594 594
FT CARBOHYD 621 621
FT CARBOHYD 672 672
FT CARBOHYD 747 747
FT CARBOHYD 823 823
FT CARBOHYD 1172 1172
FT CARBOHYD 1522 1522
FT CARBOHYD 1558 1558
SQ SEQUENCE 1905 AA; 212518 MW; 582233FA8B9AE13C CRC64;

Query Match 21.9%; Score 492; DB 1; Length 1905;
Best Local Similarity 27.9%; Pred. No. 2.7e-24;
Matches 161; Conservative 73; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYGGQIVAVADTGLDTCR-----NDS-----SMHEAFRGKITALYALGRNTNANDTNGH 68
DB 376 LRGGQILSIADTGLDGHCHCFSDSKYPIPNQVNHNRKVVY---YITHDNEYVNGH 432
QY 69 GTHVAGSVLNG-----STNKGAPQANLVFQSDGGGLGSPNLQTLFSAQYSA 121
DB 433 GTHVCGSAAGTPEDSSWAISSFSGLATDAKAFYD-LSSGSSEPTPEDYSQMKPLYDA 491
QY 122 GARTHSWGA-----AVNGATTSRNVDDYVRK-NDMTILFAAGNEGPGGTISAPGTA 176
DB 492 GARVHGSWGSVSLQGYGYGSDAGGIDAFLEYEPFSILRAAGN-NELFASLLAQATA 550
QY 177 KNATITVGATENLRPSFGS-----YADNI-----VKNRGI 199
DB 551 KNATITVGAETAHVYVSDALEYDFDGNANFORPCLPDKKYCNYYTAKCSESVNVKGL 610
QY 200 -----NHVAOFSSRGPTKGRIKPDVNAAPTFTILSARSS- 233
DB 611 QLCCPASIKQNASDFTTQPOFYNNENNGSPSSKGPHTDGLKPDIVAPGEYITSARSNG 670
QY 234 -----LAPDSFWANHDSKYAVMGSTSMATPIVAGNVAQLREHF-----VKNRGI 278
DB 671 ENSTDQCGDGL--PNANGLMSISGTSNATPLATAATILRQVLDVGYFTGSEVENKL 728
QY 279 TPKEPSLKAAIAGAADIGLY-----PKNQGWGRTLDKSLNVA 319
DB 729 LPTGSLIKALMINNAQLNGTYFWSAGSTNPSNAIFQINGANLIQGWALRWN---NWL 785
QY 320 YVNESS-----SLSTSQRAT-----YST-ATAGK 343

DB 786 YKSSNPPTPPSWIGIGLGRKQKATEWKEDSLSLGKNSCYFTYKPPSSSSSGSGGGGT 845
QY 344 P-LKISLVMSDAPATTASVTLVNDLNL-----VITAPN--GTQYVGNDFSTPYND 391
DB 846 PRIVATLVMTDPPSYSGAKFNLVNNDLILLNSDDDSIITIGNSGSLQPAKVAQP--- 902
QY 392 NWDGRNVENVFINAPOSGTGTIEVOAVNVPVGPOTFS 429
DB 903 --DTLNVVEGIIINFTKAMNYKFTIAGTNVPIGPQKFS 938
RESULT 3
SUBV_BACSU STANDARD; PRT; 806 AA.
AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
GN VPR OR IPA-45R OR BSU38090.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938892;
RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
RA Pero J.;
RT "Cloning and characterization of the gene for an additional
extracellular serine protease of Bacillus subtilis";
RL J. Bacteriol. 173:6889-6895 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020337; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hulio M.F., Ionescu M., Lubochinsky B., Marcelino L., Moser I.,
RA Presecan E., Santana M., Schneider E., Schweitzer J., Veres A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
kb region from 325 degrees to 333 degrees";
RL Mol. Microbiol. 10:371-384 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortaro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Brington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.P., Itaya M., Jones L.,
RA Joris B., Karamata D., Kishara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Kottler P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Neone D., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noona D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha S., Roche S., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M76590; AAA22881.1; --
 DR EMBL; X73124; CAAS1601.1; --
 DR EMBL; Z91123; CAB15835.1; --
 DR FIP; A41341; A41341.
 DR HSP; P00782; 2S8T.
 DR MEROPS; S08.00A; --
 DR Subtilisin; B010591; vpr.
 DR InterPro; IPR001317; PA.
 DR InterPro; IPR002029; Peptidase_S8.
 DR InterPro; IPR003020; Protease_inhib.
 DR Pfam; PF02235; PA; 1
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
 FT SIGNAL 1 28
 FT PROPEP 29 160
 FT CHAIN 161 806
 FT ACT_SITE 189 189
 FT ACT_SITE 233 233
 FT ACT_SITE 534 534
 FT ACT_SITE 806 AA; F984B3BF0B869DDD CRC64;
 SQ SEQUENCE 806 AA; 85608 MW; F984B3BF0B869DDD CRC64;
 Query Match 13.9%; Score 311.5; DB 1; Length 806;
 Best Local Similarity 23.6%; Pred. No. 5.5e-13;
 Matches 130; Conservative 56; Mismatches 157; Indels 207; Gaps 18;
 QY 18 GLYGQGI VAVADTGLDTGR-----NDSSHEAFRGKITALYALGRTN 60
 DB 177 GYTGKGIKVAIDTGVYHNHDLKKNFGQYGVDFVNDYDPKETFG-----D 225
 QY 61 NANTDNGHGHVAGSVLNGSTKGMAPQANLVFQSIMDSGGGLGLPNSLQTLFSQAYS 120
 DB 226 PRGEATDHGHVAGTVAANG-TTKGVAPDATTLLAYRVLGPGG--SGTTENVIAGVERAVQ 282
 QY 121 AGARIHNSGAANGAVGAYTTDSRNVDDYVRKNDMTILFAAGNEGNGGNTISAPGTAKNAI 180
 DB 283 DGADVMLSLGNSLNPDPWATSTAL-DWAMSEGVAVTNGSGNGPVGWTVGSGFISREAI 341
 QY 181 TVGATE-----NLRPSFGSY-----
 DB 342 SVGATQLPLNEYAVTFGYSYSSAKVMGYNKEDVDKALNKNVELVEAGIGEAQDFGKDLT 401
 QY 196 -----ADNI-----
 DB 402 GKVAIVKRGSI AFVADKADNKKAGATGMVYNNLSGEIEANVPGMVPTIKLSLEDGEKL 461
 QY 200 -----NHVAFSSRGPTKD-GRIPKDVWAPGTFILSARSSLAP 236
 DB 462 VSALKAGETKTTFKLTYSKALGQVADFSRGPVMDTWIMKPDISAPGVNIVSTIPTHDP 521
 QY 237 DSSFWNHDSEKAYMGTSWATIVAGNVAQLREHFVKRGITPKSL--LKAALIAGAA 294

RESULT 4

ID_PLS PYRPU STANDARD; PRT; 1398 AA.
 AC P72186;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Pyrolysin precursor (EC 3.4.21.-).
 GN PLS OR PF0287.
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RX MEDLINE=96355370; PubMed=8702780;
 RA Voorhorst W.G.B., Eggen R.I.L., Geerling A.C.M., Platteau C.,
 RA Siezen R.J., de Vos W.M.,
 RT "Isolation and characterization of the hyperthermostable serine
 RT protease, pyrolysin, and its gene from the hyperthermophilic archaeon
 RT *Pyrococcus furiosus*.";
 RL J. Biol. Chem. 271:20426-20431(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.,
 RT "The complete sequence of the *Pyrococcus furiosus* genome.";
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
 RX MEDLINE=21079021; PubMed=11210516;
 RA de Vos W.M., Voorhorst W.G.B., Dijkgraaf M., Kluskens L.D.,
 RA Van der Oost J., Siezen R.J.,
 RT "Purification, characterization, and molecular modeling of pyrolysin
 RT and other extracellular thermostable serine proteases from
 RT hyperthermophilic microorganisms.";
 RL Meth. Enzymol. 330:383-393(2001).
 CC -!- FUNCTION: Has endopeptidase activity toward caseins, casein
 CC fragments including alpha-SI-casein and synthetic peptides.
 CC -!- SUBCELLULAR LOCATION: Cell envelope associated.
 CC -!- PTM: LWM pyrolysin seems to be produced by autoproteolytic
 CC activation of HMW pyrolysin.
 CC -!- PTM: Glycosylated.
 CC -!- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U55835; AAB09761.1; --
 DR EMBL; AE010153; AAL80411.1; --
 DR FIP; T28159; T28159.


```

DR EMBL; AB012184; AAM40166.1; -.
DR PIR; S11890; S11890.
DR HSP; P00782; 2SBT.
DR MEROPS; S08.004; -.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00139; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 32 POTENTIAL..
FT PROPEP 33 7136 POTENTIAL..
FT CHAIN 7137 580 EXTRACELLULAR PROTEASE.
FT ACT_SITE 177 177 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 409 409 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 225 273 BY SIMILARITY.
FT DISULFID 315 352 BY SIMILARITY.
FT DISULFID 450 454 POTENTIAL.
SQ SEQUENCE 580 AA; 57228 MW; 8C9A2CAE4E7F47CB CRC64;

Query Match 12.4%; Score 278; DB 1; Length 580;
Best Local Similarity 26.0%; Pred. No. 5.5e-11;
Matches 126; Conservative 56; Mismatches 147; Indels 156; Gaps 24;

QY 21 GQGIQVAVADTGL-----DTRNDSSVHEAFPRKITALYALGRNNAND-----64
DB 168 GSGTVVAVIDTGTSHADLNANLAGYDFISDATTARDNGRDSNADEGDIYAAVECA 227
QY 65 -----TNGHGTAVAGSLVNGSTNGKMAPOA-----NLVFGSIMD 99
DB 228 GIPAASSWHGTHVAGTVAATNTTGVAGTAGAKVVPVVLKCGGSLDIADIVA 287
QY 100 SGGGLGGLPENLQ--TLFQAYAGARIHTNSWGAANGAYTTDSRVDYVRKNDMTIL 157
DB 288 SGGTVSGIPANANPAEYINNSLGGGSCSTTMQN-AINGAVSRGT-----TV 334
QY 158 FAAGNPGNGGT-SAPOTAKNAVITVGATEN--LRPSFGSYADNINHVAFSSRGPTKGR 215
DB 335 VAAGNDASVSG-SLPANCANVIAVAATTSAGKAKASYNFGTI-----377
QY 216 IKPDVNPAGPFIISARSS--LAPDSSFANHDSKIYMGTSNATPIVAGNVAQLREHFV 273
DB 378 ---DVSAPGSSILSTLNSGTTTPGS-----ASYASVNGTSMASPHVAGVALVQS--V 425
QY 274 KNRGITPK--PSLLK--AALIAGAADIGLYPNNGOGWRVTLDKSLNVAYNSS----325
DB 426 APTALTPAAVETLLKNTARALPAC-----SGCGAGIVNADAAVTAA--INGSGGGG 477
QY 326 -----SLTSQKATYGTATAGKPLKISLWSDAPASTTASVTL---VND 367
DB 478 GGGNTLTNGTPVTLGGAATGAELNYITVPAG-----SGTLTVTTSGGSD 523
QY 368 LNLVI---TAPNGTQYVNDFTSPYNDNDGRNVENVFAPNAPQSGTYIEVCAYNPVG 424
DB 524 ADLVYRAGSAPTSAYT---CRPYRS-----GNAETCTITAP-SGTYIVRLKAYS-----569
QY 425 PQTFSS 429
DB 570 --TFS 572

RESULT 6
SUBT_BAC99 ID SUBT_BAC99 STANDARD; PRT; 420 AA.
AC P28842;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Subtilisin precursor (EC 3.4.21.62).
GN SUBI.
OS Bacillus sp. (strain TA39).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=29336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92256481; PubMed=1591352;
RA Narinx E., Davail S., Feller G., Gerday C.;
RT "Nucleotide and derived amino acid sequence of the subtilisin from
the antarctic psychrotroph Bacillus TA39."
RL Biochim. Biophys. Acta 1131:111-113(1992).
CC -!- FUNCTION: Subtilisin is an extracellular alkaline serine protease,
it catalyzes the hydrolysis of proteins and peptide amides.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in P1. Hydrolyzes peptide amides.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Still active at temperatures close to 0 degrees
Celsius, it has a marked heat lability.
CC -!- MISCELLANEOUS: Secretion of subtilisin is associated with onset of
sporulation, and many mutations which block sporulation at early
stages affect expression levels of subtilisin. However, subtilisin
is not necessary for normal sporulation.
CC -!- SIMILARITY: Belongs to peptidase family S8.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X62369; CAA44227.1; -.
DR PIR; S23407; S23407.
DR HSP; Q99405; IMPT.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Metal-binding;
KW Calcium-binding; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 111 POTENTIAL.
FT CHAIN 112 420 SUBTILISIN.
FT ACT_SITE 145 145 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 360 360 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT METAL 115 115 CALCIUM (POTENTIAL).
FT METAL 154 154 CALCIUM (POTENTIAL).
SQ SEQUENCE 420 AA; 44086 MW; AB4F121BD32B26EC CRC64;

```

```

Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 4.2e-11;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GQGIQVAVADTGLDTRNDSSVHEAFPRKITAL--YALGR---NNANDTNGHGTAVGS 75
DB 136 GGGINIAVLDTGVNTN-----HPDLNNVEQCKDFVTGTTNNNSCTDRQGHGTAVGS 189
QY 76 VLNGSGTNNK---GNAPQANL-VFQSIMDSGGGLG-GLPSNLQTLFSAQYAGARIHTN-S 129
DB 190 ALADGGGTNGVYGVAPADLWAYKVLGDDSGSVADDIAAIRHAGDQATALNTKVTNMS 249
QY 130 WGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAITVGTENLR 189
DB 250 LGSSGESLITNAYN---YSNKGVLIIAAGNSGPFYQGSIGYPGALNVAVALEN-K 305

```

QY 190 PSFGSVADNINHVAQESSRGPT-KDG-----RIKPDVWAPGTILSARSLAPDSSFWAN 243
Db 306 VENGTY-----RVADFSRGYSWTGDYAIQKGDVEISAPGAAYST-----W-- 348
QY 244 HDSKIYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLYPNG 303
Db 349 FDGGYATISGTSMASPHAAGLAAKIWAQYPSASNVDRGELQVRAY---ENDILSGYYAG 405
QY 304 -----NQGWGRVTL 312
Db 406 YGDDFASGFGFATV 419
RESULT 7
AOL1 THEAQ STANDARD; PRT; 513 AA.
AC P08594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Aqualysin I precursor (EC 3.4.21.-).
GN PSTI.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RX MEDLINE=90216674; PubMed=2182621;
RA Tarada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
RT I, with NH2- and COOH-terminal pro-sequences and its processing in
RT Escherichia coli.";
RL J. Biol. Chem. 265:6576-6581 (1990).
RN [2]
RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=68235062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RA Tarada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1 and
RT characteristics of the deduced primary structure of the enzyme.";
RL Eur. J. Biochem. 173:491-497 (1988).
RN [3]
RP SEQUENCE OF 128-170.
RX MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H.,
RA Tarada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1.";
RL Eur. J. Biochem. 171:441-447 (1988).
CC -1- FUNCTION: Aqualysin I is a thermophilic alkaline serine protease.
CC the optimal temperature for its caseinolytic activity is 80
CC degrees Celsius.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DEVELOPMENTAL STAGE: Secreted from the early stationary phase
CC until the time the cells cease to grow.
CC -1- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
CC PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
CC C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
CC PROTEASES ACROSS THE OUTER MEMBRANE.
CC -1- PTM: Two disulfide bonds are present.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMEL; D90108; BAA14135.1; -.
DR EMEL; X07734; CAA30559.1; -.
DR PIR; A35742; A35742.
DR HSSP; P06873; 2PKK.
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR009020; Protease inhib.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 14
FT PROPEP 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPEP 409 513
FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6D4A50B785 CRC64;
Query Match 11.7%; Score 262.5; DB 1; Length 513;
Best Local Similarity 26.2%; Pred. No. 4.9e-10;
Matches 118; Conservative 43; Mismatches 144; Indels 145; Gaps 23;
QY 16 SYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALY-ALGRTNANDTNHGHTHVAG 74
Db 152 TVTATGRGVNVVYDTGIRT-----THREFGGRARVGYDALG--GNGQDCNGHGHVAG 203
QY 75 SVLGNSTNKGMAPQANLVFQSIIND--SGGGLGGLPSNLQTLFQAYSAGARIHTN---- 128
Db 204 TI---GGVTYGVAKAVNLVAVRVLDCNGSGTSGVIAGVDWV-----TRNHRRAVA 252
QY 129 --SWGAAVNGAYTTDSRNVDYVRKN---DMTILFAAGNEGPNGGTISAPGAKNAITVG 183
Db 253 NMSLGGGVSTA-----LDNAVKNSTAAGVVVAVAAAGNDNANACNYS-PARVAEALTVG 304
QY 184 AT--ENLRPFGSVADNINHVAQESSRGPTKGRKIPDWMAPGTILSARSLAPDSSFW 241
Db 305 ATTSDDARASFSNYGSCV-----DLFAPGASIFSA-----W 335
QY 242 ANHDSKVAYMGTSMTATPIVAGNVAQLREHFVKNRGITP---KPSLLKAALIAGAADI 298
Db 336 YTSATATQTLNGTSMATPHVAG-VAAL--YLEQNPSTPASVASAILNGATTGRLSGIGS 392
QY 299 GYPNGNQGWGRVTLQKSLINVAIVNNESSLSSTQKATYSTATAGKPLKISLVWSDAPST 358
Db 393 GSPN-----RLYSLILSSGS-----GSTAPCTS 415
QY 359 ---TASVTLVNDLNLVITAPNGTOY-----VGNDFTSFYNDNWDGRN--- 397
Db 416 CSYVTGSLSGPDYNF---QPNGYIYSPAGTHPAWLRGPGATDP-DLYLWRWDGSRWLT 471
QY 398 -----NVENVFINAPQSGTYTIEVOAYN 420
Db 472 VGSSTGPTSEBSLSYSGTAGYLLRWIYAYS 501
RESULT 8
WRA_BACSU
ID WRA_BACSU STANDARD; PRT; 894 AA.
AC P54423; O06726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Cell wall-associated protease precursor (EC 3.4.21.-) [Contains: Cell
GN WRA OR BSU10770.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.
RC STRAIN=168;
RX MEDLINE=57159234; PubMed=9004506;
RA Margot P., Karamata D.;
RT "The wprA gene of *Bacillus subtilis* 168, expressed during exponential
growth, encodes a cell-wall-associated protease.";
RL Microbiology 142:3437-3444(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=168;
RC STRAIN=168;
RX MEDLINE=98015415; PubMed=933931;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289
degrees) in *Bacillus subtilis*.";
RL Microbiology 143:3305-3308(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=168;
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
Hilbert B., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Pario V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadia Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
Sakakini A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
Tateuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Yata K.,
Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOLYCAN
DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
CC -!- SUBCELLULAR LOCATION: Cell-wall bound.
CC -!- PTM: PROCESSED INTO CWBP23 AND CWBP52.
CC -!- SIMILARITY: Belongs to peptidase family 58.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U58981; AAC25926.1; -;
CC EMBL; Y09476; CAA70641.1; -;
CC EMBL; Z98109; CAB12917.1; -;
CC PIR; F69730; F69730.
CC HSP; Q45670; LDI1.
CC MEROPS; S08.004; -;
CC Subtilisin; BG11846; wprA.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 31
FT CHAIN 32 894
FT CHAIN 32 ?
FT PROPEP 7 413
FT CHAIN 414 894
FT ACT_SITE 466 466
FT ACT_SITE 497 497
FT ACT_SITE 650 650
FT CONFLICT 9 9
FT CONFLICT 14 14
FT CONFLICT L -> I (IN REF. 1).
SQ SEQUENCE 894 AA; 96487 MW; 0F67C35E55F8DBC CRC64;
Query Match 11.3%; Score 253.5; DB 1; Length 894;
Best Local Similarity 24.7%; Pred. No. 3.9e-09;
Matches 108; Conservative 60; Mismatches 141; Indels 129; Gaps 19;
QY 25 IVAVADTGLDTRNDSSMHEAFRGKITALYA---LGRTNANDTNGHGHVAGSVLG--- 78
DB 457 LIAVVDITGVSTLAD-----LKGKVRTDLGHNFVGRNMMDDQGHGTHVAGIIAAQSD 510
QY 79 NGSTNKGMAQANLVFQSIWDSGGGLGSLNSLTQTFQSAQYAGARIHTNSWGAAYNGAY 138
DB 511 NGYSMTGLNAKAKIIPVKVLDLSAG--SGDTEQALGKIYAADKGVKINISLG-----GGY 564
QY 139 TTDSRNVDVVRKNDMILFAAGNEGPNGGGTISAPGAKNAITVGATENLRPSFGSYADN 198
DB 565 SRVLEFALKVAADKNVLIAAAGNDNGEN--ALSYPASSKYVMSVGT-----NR 611
QY 199 INHVAQSSRGRTKDGRIKPDVMAPTFISASSLAPDSFWANHDSKYAYMGTSMAT 258
DB 612 MDMTADFSTNYGKGL-----DISAPGSDI-----PSLVPNGN-----VTYMSGTSMAT 653
QY 259 PIVAGNVAQRLREHFVKNGRI--TPKPSLLKAALIAAGADICGLGYPNGNQ----- 305
DB 654 PYAAAAAGLL---FAQPKLKRTEVEMDKKT-----ADDISFESVDGSEBELYDDYDPI 706
QY 306 -----GWRVTLDKSLNVA-----YVNESSLSSTQKATYS----- 336
DB 707 EIPKTPGVDMHSGYGRNLNWKVSAADLQKVKNKLESTQTAVRGSAKEGTLIEVMNGKKK 766
QY 337 -FTATAGK--PLKISLVMSDAPASTTASVTIVNDLNLVITAPNGTQVGVNDFTSPYNDNW 393
DB 767 LGSAGAGKDNKAFKNI-----ATQKQDQVLYLKATG----- 798
QY 394 DGRNNVENVFINAPQSGT 411
DB 799 DAKTSYKVVVVKPKPSGT 816
RESULT 9
ALP_CBPAC
ID ALP_CBPAC STANDARD; PRT; 402 AA.
AC P29118;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN ALP.
OS Cephalosporium acremonium (Acremonium chrysogenum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;
OC Acremonium.
OX NCBI_TaxID=5044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9129283; PubMed=1368696;
RA Isogai T., Fukagawa M., Kojo H., Kohsaka M., Aoki H., Imanaka H.;

```

RT "Cloning and nucleotide sequences of the complementary and genomic
RT DNAs for the alkaline protease from Acremonium chrysogenum.";
RL Agric. Biol. Chem. 55:471-477(1991).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00923; BA00765.1; -.
DR PIR; J00332; J00332.
DR HSP; P06873; 2PKK.
DR MEROPS; S08.UFA; -.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
KW SIGNAL 1 20
DR FT PROPEP 21 120 POTENTIAL.
DR FT CHAIN 121 402 ALKALINE PROTEINASE
DR FT ACT_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR SQ SEQUENCE 402 AA; 42099 MW; 8D030CCD42D918E1 CRC64;
Query Match 11.0%; Score 246.5; DB 1; Length 402;
Best Local Similarity 30.3%; Pred No. 3.9e-09;
Matches 91; Conservative 32; Mismatches 102; Indels 75; Gaps 14;
QY 21 GGGQIVAVADTGLDTRNDSSMHEAFRGK-ITALYALGRTNANDTNGHGHVAGSVLGN 79
DB 151 GSGTYAVYVDVTGI-----LESHNEFSGRAITGYNVAGGSGN--ADTNGHGHVAGTI--- 199
QY 80 GSTNKGMAPOANLYFQSIIMDSGGG-----LGGIPSLNLTLPSCAYSAGARI-----HT 127
DB 200 GGRTYGVAKNNLTIAVKVFRGSSSTSIILDGFNVAWINDIIRGQNKAALISMILGGGYS 259
QY 128 NSWGAAVANGAYTTDSRNVDDTVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVG--AT 185
DB 260 SAFNNAVNTAY--SRGVLSVV-----AAGNDNQNAAYS-PASAANAITVGSITAS 306
QY 186 ENLRPFGSYADININHAQFSRGPDTKGRIKPDVWAPGFTILSARSLAPDSSFWANHD 245
DB 307 NWARSSFNYSGLV-----DIFAPGTSILSA-----WICGN 337
QY 246 SKYAVYGGTSMATPIVAGNVQAQLREHFVKNRGITPKPSLLKAALIAAGAADI---GLGYPN 302
DB 338 SATNIISGTSNATPVTVGVLYLQ-----ALEGLTSCAAARLNALATTGRVSNPGSGSPN 393
RESULT 10
PROA_VIBAL
ID PROA_VIBAL STANDARD; PRT; 534 AA.
AC P16588;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline serine exoprotease A precursor (EC 3.4.21.-).
DE PROA.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
CC NCBI_TaxID=663;
CC [1]
RP SEQUENCE FROM N.A.

```

RESULT 11
 THES_BACSP STANDARD; PRT; 401 AA.
 AC Q45670; 2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Thermophilic serine proteinase precursor (EC 3.4.21.-) (Ak.1
 protease).
 OS Bacillus sp. (strain AK1).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1409;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=95085262; PubMed=7993087;
 RA Maciver B., McHale R.H., Saul D.J., Bergquist P.L.;
 RT "Cloning and sequencing of a serine proteinase gene from a
 thermophilic Bacillus species and its expression in Escherichia
 coli";
 RL Appl. Environ. Microbiol. 60:3981-3988(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 122-401.
 RX MEDLINE=20057863; PubMed=10588904;
 RA Smith C.A., Toogood H.S., Baker H.M., Daniel R.M., Baker E.N.;
 RT "Calcium-mediated thermostability in the subtilisin superfamily: the
 crystal structure of Bacillus Ak.1 protease at 1.8-A resolution.";
 RL J. Mol. Biol. 294:1027-1040(1999).
 CC -I- COFACTOR: Binds 3 calcium ions and 1 sodium ion per subunit.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- MISCELLANEOUS: Has a pH optimum of 8.5, a temperature optimum of
 75 degrees Celsius.
 CC -I- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L29506; AAA63688.1; -
 DR PIR; I39974; 139974.
 DR PDB; 1DBI; 18-NOV-99.
 DR MEROPS; S08.009; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
 KW Signal; 3D-structure.
 KM Signal; 24 POTENTIAL.
 FT SIGNAL 1 24
 FT PROPEP 25 121
 FT CHAIN 122 401 THERMOPHILIC SERINE PROTEINASE.
 FT ACT_SITE 160 160 CHARGE RELAY SYSTEM.
 FT ACT_SITE 193 193 CHARGE RELAY SYSTEM.
 FT ACT_SITE 347 347 CHARGE RELAY SYSTEM.
 FT METAL 126 126 CALCIUM 1.
 FT METAL 168 168 CALCIUM 3. (VIA CARBONYL OXYGEN).
 FT METAL 169 169 CALCIUM 1.
 FT METAL 171 171 CALCIUM 3.
 FT METAL 179 179 CALCIUM 2.
 FT METAL 184 184 CALCIUM 2.
 FT METAL 186 186 CALCIUM 2. (VIA CARBONYL OXYGEN).
 FT METAL 204 204 CALCIUM 1.
 FT METAL 204 204 CALCIUM 3.
 FT METAL 207 207 CALCIUM 1.
 FT METAL 209 209 CALCIUM 1. (VIA CARBONYL OXYGEN).
 FT METAL 211 211 CALCIUM 1. (VIA CARBONYL OXYGEN).
 FT METAL 211

297 297 METAL 10.8%; Score 242.5; DB 1; Length 401;
 300 300 FT METAL 30.9%; Pred. No. 7.1e-09;
 323 323 FT METAL 30; Mismatches 100; Indels 51; Gaps 10;
 258 258 FT DISULFID
 127 127 FT TURN
 129 129 FT HELIX
 132 132 FT TURN
 135 135 FT TURN
 136 136 FT HELIX
 137 137 FT TURN
 140 140 FT TURN
 140 140 FT TURN
 141 141 FT HELIX
 144 144 FT TURN
 145 145 FT TURN
 152 152 FT TURN
 153 153 FT STRAND
 155 155 FT TURN
 166 166 FT TURN
 168 168 FT TURN
 173 173 FT TURN
 174 174 FT STRAND
 179 179 FT STRAND
 183 183 FT TURN
 184 184 FT STRAND
 184 184 FT STRAND
 193 193 FT HELIX
 202 202 FT TURN
 219 219 FT STRAND
 228 228 FT TURN
 229 229 FT HELIX
 234 234 FT TURN
 246 246 FT TURN
 247 247 FT STRAND
 251 251 FT STRAND
 263 263 FT HELIX
 274 274 FT TURN
 275 275 FT TURN
 278 278 FT STRAND
 285 285 FT STRAND
 288 288 FT TURN
 298 298 FT TURN
 301 301 FT STRAND
 306 306 FT TURN
 308 308 FT TURN
 312 312 FT STRAND
 312 312 FT TURN
 314 314 FT TURN
 315 315 FT STRAND
 316 316 FT TURN
 320 320 FT STRAND
 321 321 FT STRAND
 324 324 FT TURN
 332 332 FT STRAND
 333 333 FT TURN
 336 336 FT TURN
 338 338 FT STRAND
 339 339 FT STRAND
 343 343 FT HELIX
 346 346 FT TURN
 363 363 FT TURN
 364 364 FT HELIX
 376 376 FT TURN
 377 377 FT STRAND
 379 379 FT TURN
 382 382 FT STRAND
 383 383 FT TURN
 384 384 FT TURN
 385 385 FT STRAND
 387 387 FT TURN
 388 388 FT STRAND
 391 391 FT TURN
 392 392 FT HELIX
 395 395 FT TURN
 400 400 FT TURN
 401 401 SQ SEQUENCE 401 AA; 42835 MW; 1C736EF4A89F256F CRC64;
 Query Match 10.8%; Score 242.5; DB 1; Length 401;
 Best Local Similarity 30.9%; Pred. No. 7.1e-09;
 Matches 81; Conservative 30; Mismatches 100; Indels 51; Gaps 10;
 QY 11 DVAQSSYGLYGGQGVAVADTGLDGTGRNDSMHEAPRGKITALLY-ALGRNNANDTNGHG 69
 DB 146 DVTKGS-----SGQEIAVIDTGVD-----YTHPDLGKVIKGYDFVDNDYDPMDLNHHG 194
 QY 70 THVAG---SVLGNSTNKGWAPQANLVFQSIMDSGGGLGSLPSNLQTLFSQAYSAGARTH 126
 DB 195 THVAGIAAETNNATGIAGWPNTRILAVPALDNG--SGTSLDIADAIYAADSAREVI 252
 QY 127 TNSWGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGAKNAITYGATE 186
 DB 253 NLSLGC---DCHTTTLENVAVNKNKGSVVVAAAGNNG--SSTTFEPASVENYAVGA-- 305
 QY 187 NLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAPGTFILSARSLAPDSSFWANHDS 246
 DB 306 -----VDQYDRLASFSNYGTW-----VDVAPGVDIVSTITG-----N 338

QY 247 KYAWGCTSMATPIVAGNVAQL 268
 Db 339 RYAYMGTSMAHPVAGLAALL 360

RESULT 12
 SUBS_BAGLE STANDARD; PRT; 269 AA.
 ID SUBS_BAGLE STANDARD; PRT; 269 AA.
 AC P29600;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Subtilisin Savinase (EC 3.4.21.62) (Alkaline protease).
 OS Bacillus lentus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1467;
 RN [1]
 RN X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
 RP MEDLINE=92148829; PubMed=1738156;
 RA Betzel C., Klupsch S., Papendorf G., Hastrup S., Branner S.,
 RA Wilson K.S.;
 RT "Crystal structure of the alkaline proteinase Savinase from *Bacillus*
 RT lentus at 1.4-A resolution.";
 RL J. Mol. Biol. 223:427-445(1992).
 RN [2]
 RN STRUCTURE BY NMR.
 RP MEDLINE=96184541; PubMed=8654411;
 RA Remerowski M.L., Pepermans H.A.M., Hilbers C.W., van de Ven F.J.M.;
 RA "Backbone dynamics of the 269-residue protease Savinase determined
 RT from 15N-NMR relaxation measurements.";
 RL Eur. J. Biochem. 235:629-640(1996).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (0.78 ANGSTROMS).
 RP MEDLINE=98426039; PubMed=9753430;
 RA Kuhn P., Knapp M., Soltis S.M., Ganshaw G., Thoenes M., Bott R.;
 RA "The 0.78-A structure of a serine protease: *Bacillus lentus*
 RT subtilisin.";
 RL Biochemistry 37:13446-13452(1998).
 CC -!- FUNCTION: Subtilisin is an extracellular alkaline serine protease,
 CC it catalyzes the hydrolysis of proteins and peptide amides.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in P1. Hydrolyzes peptide amides.
 CC -!- COFACTOR: Binds 2 calcium ions per subunit.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
 CC Savinase by Novozymes.
 CC -!- MISCELLANEOUS: Secretion of subtilisin is associated with onset of
 CC sporulation, and many mutations which block sporulation at early
 CC stages affect expression levels of subtilisin. However, subtilisin
 CC is not necessary for normal sporulation.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 DR PDB; 1C9J; 06-OCT-99.
 DR PDB; 1C9M; 10-JAN-01.
 DR PDB; 1C9N; 10-JAN-01.
 DR PDB; 1GCI; 11-NOV-98.
 DR PDB; 1I4V; 11-JUL-01.
 DR PDB; 1JEA; 26-NOV-97.
 DR PDB; 1SVN; 14-OCT-96.
 DR MEROPS; S08.103; --
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILISIN ASP; 1.
 DR PROSITE; PS00137; SUBTILISIN HIS; 1.
 DR PROSITE; PS00138; SUBTILISIN SER; 1.
 DR PROSITE; PS00139; SUBTILISIN ASP; 1.
 KW Hydrolase; Sporulation; Serine protease; Metal-binding;
 Calcium-binding; 3D-structure.
 FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
 FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
 FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
 FT ACT_SITE 2 2 CALCIUM 1.
 FT METAL 40 40 CALCIUM 1.

FT METAL 73 73 METAL
 FT METAL 75 75 METAL
 FT METAL 77 77 METAL
 FT METAL 79 79 METAL
 FT METAL 163 163 METAL
 FT METAL 165 165 METAL
 FT METAL 168 168 METAL
 FT STRAND 2 2
 FT HELIX 6 10
 FT TURN 11 12
 FT HELIX 13 18
 FT TURN 19 20
 FT TURN 24 25
 FT STRAND 27 32
 FT TURN 39 40
 FT STRAND 43 48
 FT TURN 51 52
 FT HELIX 62 71
 FT STRAND 78 78
 FT TURN 84 85
 FT STRAND 87 92
 FT TURN 94 94
 FT TURN 96 97
 FT STRAND 100 100
 FT HELIX 102 114
 FT TURN 115 116
 FT STRAND 119 122
 FT STRAND 126 126
 FT HELIX 131 142
 FT TURN 143 144
 FT STRAND 146 150
 FT STRAND 161 161
 FT TURN 162 164
 FT TURN 166 167
 FT STRAND 168 174
 FT TURN 176 177
 FT STRAND 180 180
 FT TURN 182 183
 FT TURN 188 189
 FT STRAND 190 195
 FT STRAND 199 203
 FT TURN 204 206
 FT STRAND 207 211
 FT HELIX 214 231
 FT TURN 233 234
 FT HELIX 237 246
 FT TURN 247 247
 FT STRAND 249 249
 FT HELIX 254 257
 FT TURN 258 259
 FT STRAND 261 261
 FT HELIX 264 267
 FT TURN 268 269
 SQ SEQUENCE 269 AA; 25698 MW; 4D89F8779999BF8D CRC64;

Query Match 10.7%; Score 240; DB 1; Length 269;
 Best Local Similarity 31.1%; Pred. No. 6.2e-09;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY 8 VKADVAQSSYGLYGQGIIVAVADTGLDTGRDSSNHEAFRGKITALYALGRNNANDTNG 67
 Db 11 VOAPAAHNR-GLTSGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDNG 61
 QY 68 HGTHVAGSV--LNGSTNKGMAPOANLVFQSTIMDSGGGLGGLPSNLQTLFSQAVSAGARI 125
 Db 62 HGTHVAGTIALNLSIGVLGVAPSRLYAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 119
 QY 126 HTNSWGA-----AVNGAYTDSRVDDYVRKNKNTILFAAGNEGPNGTISAPGTA 176
 Db 120 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165
 QY 177 KNAITVGATE--NLRPSFGSYADNHNHVAQFSRGPTKDGRIKPDVMAPGTIFLSARSSL 234

Db 166 ANAMAVGATDQNNRASFQYAGL-----DIVAPGVNVQSTVPG- 205

QY 235 APDSFVWANHDSKYAYMGCTSMATPVGNA-----QLRHHFVKN 275

Db 206 -----STYASLNGTSMATPVAGAAALVQKNPSNVQIRNH-LKN 246

RESULT 13

ELYA_BACAO

ID ELYA_BACAO STANDARD; PRT; 380 AA.

AC P27693;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alkaline protease precursor (EC 3.4.21.-)

OS Bacillus alcalophilus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1445;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PB92;

RC MEDLINE=91282483; PubMed=2059048;

EX van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoek R.A.,

RA Quax W.J.;

RT "Cloning, characterization, and multiple chromosomal integration of a

RT Bacillus alkaline protease gene.";

RL Appl. Environ. Microbiol. 57:901-909(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).

RC STRAIN=PB92;

RC MEDLINE=92390330; PubMed=1518788;

RA van der Laan J.C., Teplyakov A.V., Keiders H., Kalk K.H., Misset O.,

RA Mulleners L.J.M., Dijkstra B.W.;

RT "Crystal structure of the high-alkaline serine protease PB92 from

RT Bacillus alcalophilus.";

RL Protein Eng. 5:405-411(1992).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).

RC MEDLINE=93078250; PubMed=1447775;

RA Sobek H., Hecht H.-J., Ashle W., Schonburg D.;

RT "X-ray structure determination and comparison of two crystal forms of

RT a variant (Asn15Arg) of the alkaline protease from Bacillus

RT alcalophilus refined at 1.85-A resolution.";

RL J. Mol. Biol. 228:108-117(1992).

RN [4]

RP STRUCTURE BY NMR OF 112-380.

RC STRAIN=PB92;

RC MEDLINE=97277237; PubMed=9115441;

RA Martin J.R., Mulder F.A., Katimi-Nejad Y., van der Zwan J.,

RA Mariani M., Schipper D., Boelens R.;

RT "The solution structure of serine protease PB92 from Bacillus

RT alcalophilus presents a rigid fold with a flexible substrate-binding

RT site.";

RL Structure 5:521-532(1997).

CC -|- COFACTOR: Binds 2 calcium ions per subunit.

CC -|- SUBCELLULAR LOCATION: Secreted.

CC -|- SIMILARITY: Belongs to peptidase family S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M65086; AAA22212.1; --

DR EMBL; A13738; CAA01128.1; --

DR PIR; A49778; A49778.

DR PDB; 1AH2; 15-APR-98.

DR MEROPS; S08.038; --

DR InterPro; IPR000209; Peptidase S8.

DR InterPro; IPR009020; Protease_inhib.

DR Pfam: PF00082; Peptidase S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;

KW Signal; 3D-structure.

FT SIGNAL 1 27 POTENTIAL.

FT PROPEP 28 112

FT CHAIN 113 380 ALKALINE PROTEASE.

FT ACT_SITE 143 143 CHARGE RELAY SYSTEM.

FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.

FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.

FT METAL 113 113 CALCIUM 1.

FT METAL 151 151 CALCIUM 1.

FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 186 186 CALCIUM 1.

FT METAL 188 186 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 190 190 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 274 274 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT HELIX 117 120

FT TURN 121 123

FT HELIX 125 130

FT TURN 131 131

FT TURN 135 136

FT STRAND 138 142

FT TURN 150 151

FT STRAND 156 157

FT HELIX 173 182

FT STRAND 198 201

FT TURN 207 208

FT HELIX 213 226

FT STRAND 230 233

FT HELIX 242 252

FT TURN 253 255

FT STRAND 257 258

FT TURN 267 268

FT TURN 272 275

FT STRAND 279 279

FT STRAND 282 285

FT STRAND 291 291

FT TURN 299 300

FT STRAND 303 306

FT TURN 310 314

FT TURN 315 317

FT STRAND 318 322

FT HELIX 325 342

FT TURN 344 345

FT HELIX 348 358

FT HELIX 375 380

SQ SEQUENCE 380 AA; 38853 MW; 539EA72771B6682C CRC64;

Query Match 10.7%; Score 240; DB 1; Length 380;

Best Local Similarity 31.1%; Pred. No. 9.7e-09;

Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY 8 VKADVAOSSGLYGGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNNDNTNG 67

Db 122 VQPAAHNR-GLTSGVKVAVLDIGIST-----HPDLNIRGGASFVPGEP-STQDNG 172

QY 68 HGTHVAGSV--LNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLOTLPQAYSAGARI 125

Db 173 HGTHVAGTIAALNNSIGVLGVAPNAELYAVKVLGASG--SGSVSSIAQGLEWAGNGMHV 230

QY 126 HTNSKGA-----AVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTISAPCTA 176

Db 231 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYARY 276

QY 177 KNATTVGATE--NLRPSFGSYADNINHVAQPSRGPTKDKRIKPDVMAPTFILSARSSL 234

Db 277 ANAMAVGATDQNNRASFQYAGL-----DIVAPGVNVQSTVPG- 316


```
QY 235 APDSFWANHDSKYAYMGGTSMATPIVAGNVA-----OLREHFVK 275
Db 317 -----SYASLNGTSMATPHVAGAAALVKQKNSWENVQIRNH-LKN 357

RESULT 14
ELVA_BACCS STANDARD; PRT; 380 AA.
AC P41362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE=93043753; PubMed=1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
RA Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
RT structural gene for alkaline serine protease from alkaliphilic
RT Bacillus sp. 221.";
RL Biosci. Biotechnol. Biochem. 56:1455-1460 (1992).
RN [2]
RP SEQUENCE OF 112-129.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA Horikoshi K.;
RL (In) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
RL Weinheim (1991);
CC -|- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; S48754; AAC60420.1; -
DR EMBL; D13157; BAA02442.1; -
DR EMBL; A26817; CAA01836.1; -
DR EMBL; A22550; CAA01611.1; -
DR HSSP; P29600; LGCI.
DR MEROPS; S08.103; -.
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR003020; Protease Inhib.
DR Pfam; PF00062; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR KX Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
KX Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 111
FT CHAIN 112 380
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT METAL 113 113 CALCIUM 1 (BY SIMILARITY).
FT METAL 151 151 CALCIUM 1 (BY SIMILARITY).
FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN)
FT METAL 186 186 (BY SIMILARITY).
FT METAL 188 188 CALCIUM 1 (BY SIMILARITY).
FT METAL 186 186 CALCIUM 1 (VIA CARBONYL OXYGEN)
```

```
FT METAL 190 190 (BY SIMILARITY).
FT METAL 274 274 CALCIUM 1 (VIA CARBONYL OXYGEN)
FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN)
SQ SEQUENCE 380 AA; 38926 MW; 5F73ABC6D5B6831 CRC64;

Query Match 10.7%; Score 240; DB 1; Length 380;
Best Local Similarity 31.1%; Pred. No. 9.7e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY 8 VKADVAQSSYGLYGGQGIIVADVDTGDTGRNDSMHEAFKGTALYALORTNNANTNG 67
Db 122 VQAPAAHNR-GLTSGVKVAVLDGTIST-----HPDLNIRGASFVGPPE-STQDNG 172

QY 68 HGTHVAGSV--LNGSTNKGMAPQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYSAGARI 125
Db 173 HGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASG--SGSVSSIAQGLEWAGNGMHV 230

QY 126 HTNSWGA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNPGNGGTISAPGTA 176
Db 231 ANLSIGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276

QY 177 KVATVGAATE--NLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVAVAGTFFILSARSSL 234
Db 277 ANAMAVGATDQNNRASFSQYAGL-----DIVAPGVNVQSTYFG- 316

QY 235 APDSFWANHDSKYAYMGGTSMATPIVAGNVA-----OLREHFVK 275
Db 317 -----SYASLNGTSMATPHVAGAAALVKQKNSWENVQIRNH-LKN 357

RESULT 15
PRTM_BACSP STANDARD; PRT; 269 AA.
AC Q99405;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE M-protease (EC 3.4.21.-).
OS Bacillus sp. (strain KSM-K16).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
RA Kobayashi T., Ito S., Yamashita O.;
RT "Structure of a new alkaline serine protease (M-protease) from
RT Bacillus sp. KSM-K16."
RL Acta Crystallogr. D 51:199-206 (1995).
RN [2]
RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
RX MEDLINE=93558832; PubMed=7632397;
RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
RA Koike K., Kawai S., Ito S.;
RT "Purification and properties of an alkaline protease from
RT alkalophilic Bacillus sp. KSM-K16."
RL Appl. Microbiol. Biotechnol. 43:473-481 (1995).
CC -|- COFACTOR: Binds 2 calcium ions per subunit.
CC -|- SUBUNIT: Monomer.
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC PDB; IMPT; 22-JUN-94.
DR MEROPS; S08.010; -.
DR InterPro; IPR000209; Peptidase S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
```

KW Hydrolase; Serine protease; Metal-binding; Calcium-binding;
KW 3D-structure.
FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
FT METAL 2 2 CALCIUM 1.
FT METAL 40 40 CALCIUM 1.
FT METAL 73 73 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 75 75 CALCIUM 1.
FT METAL 77 77 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 79 79 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 163 163 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 165 165 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 6 10
FT TURN 11 12
FT HELIX 13 18
FT TURN 19 20
FT TURN 24 25
FT TURN 27 32
FT TURN 39 40
FT STRAND 43 48
FT TURN 51 52
FT HELIX 62 71
FT TURN 84 85
FT STRAND 87 92
FT TURN 96 97
FT HELIX 102 114
FT TURN 115 116
FT STRAND 119 122
FT STRAND 126 126
FT HELIX 131 142
FT TURN 143 144
FT STRAND 146 150
FT STRAND 161 161
FT TURN 162 164
FT TURN 166 167
FT STRAND 169 174
FT TURN 176 177
FT STRAND 180 180
FT TURN 182 183
FT TURN 188 189
FT STRAND 192 195
FT STRAND 199 203
FT TURN 204 206
FT STRAND 207 211
FT TURN 214 231
FT HELIX 233 234
FT TURN 237 246
FT TURN 247 247
FT STRAND 249 249
FT HELIX 254 257
FT TURN 258 259
FT STRAND 261 261
FT HELIX 264 267
FT TURN 268 269
SQ SEQUENCE 269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match 10.6%; Score 239; DB 1; Length 269;
Best Local Similarity 31.1%; Pred. No. 7.2e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
QY 8 VKADVAGSYGLYQGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRTNNDNTNG 67
DB 11 VQAPAHNR-GLTSGVGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDNG 61
QY 68 HGTHVAGSV--LQNGSTNKGMAFQANLVFQSIQSGGLGGLPSNLQTLFQAYSAQARI 125
DB 62 HGTHVAGTIAALNNSIGLVGAPSAELIYAVKVLGASG--SGSVSSIAQGLEWAGNGMHV 119
QY 126 HTNSWGA-----AVNGAYTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTA 176
DB 120 ANLSLGSFSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165

QY 177 KNAITVIGATE--NLRPSFGSYADNINHVAFSSRGFTKGRIKPDVMAPGTILSARSSL 234
DB 166 ANAMAVGATDQNNNRASFSQYGAGL-----DIVAPGVNVQSTYPG- 205
QY 235 APDSSFWANHDSKYAYMGGTSMATPIVAGNVA-----QLRHFVKN 275
DB 206 -----STYASLNGTSMATPHVAGVAALVKQKPSWNSNVOIRNH-LKN 246

Search completed: March 10, 2004, 14:53:39
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:50:43 ; Search time 45 Seconds
(without alignments)
3042.998 Million cell updates/sec

Title: US-09-985-689A-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVPGPQTFLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_muc:**
- 8: sp_organelle:**
- 9: sp_phage:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriaph:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2242	99.8	640	Q3UV9	Q3UV9 bacillus sp
2	2178	96.9	639	Q3UV9	Q3UV9 bacillus sp
3	2138	95.1	434	Q3UV9	Q3UV9 bacillus sp
4	1993.5	88.7	433	Q3UV9	Q3UV9 bacillus sp
5	1989.5	88.5	433	Q3UV9	Q3UV9 bacillus sp
6	1982.5	88.2	433	Q3UV9	Q3UV9 bacillus sp
7	506.5	22.5	1825	Q3UV9	Q3UV9 dictyosteli
8	442	19.7	1702	Q3UV9	Q3UV9 dictyosteli
9	409	18.2	654	Q3UV9	Q3UV9 dictyosteli
10	401.5	17.9	561	Q3UV9	Q3UV9 dictyosteli
11	399	17.8	1239	Q3UV9	Q3UV9 dictyosteli
12	381	17.0	430	Q3UV9	Q3UV9 dictyosteli
13	371	16.5	1253	Q3UV9	Q3UV9 dictyosteli
14	366	16.3	1102	Q3UV9	Q3UV9 dictyosteli
15	363.5	16.2	1208	Q3UV9	Q3UV9 dictyosteli
16	349	15.5	1245	Q3UV9	Q3UV9 dictyosteli

17	348	15.5	1105	2	Q8KH6
18	348	15.5	1139	16	Q8KH6
19	347.5	15.5	444	16	Q8KH6
20	338	15.0	412	2	Q8KH6
21	337	15.0	412	16	Q8KH6
22	336	15.0	1237	2	Q8KH6
23	329.5	14.7	824	2	Q45464
24	329.5	14.7	1220	16	Q9L0A0
25	327.5	14.6	435	16	Q8KH6
26	321.5	14.3	891	1	Q3635
27	320.5	14.3	442	16	Q3635
28	308.5	13.7	431	2	Q8KH6
29	307.5	13.7	1398	1	Q8KH6
30	294.5	13.1	434	2	Q54327
31	292	13.0	1135	1	Q8KH6
32	282.5	12.6	586	16	Q8KH6
33	281	12.5	419	2	Q45681
34	280.5	12.5	959	16	Q8KH6
35	280	12.5	799	16	Q8KH6
36	279.5	12.4	1345	1	Q54437
37	275.5	12.3	621	2	Q9F486
38	275.5	12.3	621	2	Q53401
39	275	12.2	1101	16	Q82CF0
40	272.5	12.1	568	16	Q8PMCO
41	268.5	11.9	403	2	Q45463
42	268.5	11.9	814	16	Q82VB3
43	267.5	11.9	575	16	Q8PNW1
44	266	11.8	1098	16	Q9L128
45	265.5	11.8	966	16	Q8PB28

ALIGNMENTS

RESULT 1

Q3UV9	PRELIMINARY;	PRT;	640 AA.
ID	Q3UV9		
AC	Q3UV9		
DT	01-DEC-2001 (TREMREL. 19, Created)		
DT	01-OCT-2002 (TREMREL. 22, Last sequence update)		
DT	01-JUN-2003 (TREMREL. 24, Last annotation update)		
DE	Protease.		
GN	PROF.		
OS	Bacillus sp. KSM-KP43.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxID=109322;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KP43;		
RA	Itch S., Saeki K.;		
RT	"new protease";		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB051423; BAB55674.2; -		
DR	GO; GO:0004289; F:subtilisin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000209; Peptidase_S8.		
DR	InterPro; IPR007280; PPC.		
DR	Pfam; PF00082; Peptidase_S8; 1.		
DR	Pfam; PF04151; PPC; 1.		
DR	PRINTS; PR00723; SUBTILISIN		
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.		
DR	PROSITE; PS00138; SUBTILASE_SER; 1.		
SQ	SEQUENCE 640 AA; 67991 MW; 4BBAF77E9D592C15 CRC64;		

Query Match 99.8%; Score 2242; DB 2; Length 640;
Best Local Similarity 99.8%; Pred. No. 6.9e-122;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db	207	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 266
QY	61	NANDTNGHGTAVAGSVLNGSTKMGAPQANLVFQISIMDSGGGLGLPSNLQTLFQSAYS 120

```

Db 267 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N I Q T L F S Q A F S 326
QY 121 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 180
Db 327 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 386
QY 181 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 240
Db 387 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 446
QY 241 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 300
Db 447 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 506
QY 301 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 360
Db 507 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 566
QY 361 S V T L V N D L N I V T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P Q S G T Y T I E V Q A Y N 420
Db 567 S V T L V N D L N I V T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P Q S G T Y T I E V Q A Y N 626
QY 421 V P V G P Q T F S L A I V N 434
Db 627 V P V G P Q T F S L A I V N 640

```

RESULT 2

```

Q9AQR3 ID Q9AQR3 PRELIMINARY; PRT; 639 AA.
AC Q9AQR3
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Protease.
GN PROA.
OS Bacillus sp. 9860.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133778;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9860;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319 (2000).
DR EMBL; AB046403; BAB21266.2; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0004289; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 639 AA; 68185 MW; 316AF6FDBE4FF54 CRC64;

```

```

Query Match 96.1%; Score 2178; DB 2; Length 639;
Best Local Similarity 96.1%; Pred. No. 3.5e-118;
Matches 417; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
Db 206 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 265
QY 61 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N I Q T L F S Q A F S 120

```

```

Db 266 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N I Q T L F S Q A F S 325
QY 121 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 180
Db 326 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 386
QY 181 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 240
Db 386 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 445
QY 241 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 300
Db 446 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 505
QY 301 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 360
Db 506 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 565
QY 361 S V T L V N D L N I V T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P Q S G T Y T I E V Q A Y N 420
Db 566 S V T L V N D L N I V T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P Q S G T Y T I E V Q A Y N 625
QY 421 V P V G P Q T F S L A I V N 434
Db 626 V P V G P Q T F S L A I V N 639

```

RESULT 3

```

Q9AQR0 ID Q9AQR0 PRELIMINARY; PRT; 434 AA.
AC Q9AQR0
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROB.
OS Bacillus sp. NV1.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NV1;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319 (2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046406; BAB21269.1; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 434
SQ SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;

```

```

Query Match 95.1%; Score 2138; DB 2; Length 434;
Best Local Similarity 93.3%; Pred. No. 4.4e-116;
Matches 405; Conservative 20; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60

```

Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
QY 61 NANTDNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSSGGLGSLPSNLQTLFSQAYS 120
Db 61 NANTDNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSSGGLGSLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAFGPTFILSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAFGPTFILSARSLAPDSSF 240
QY 241 WANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAQVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAQVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNMDGRNVNENVFINAPQSGTYYTIEVOAYN 420
Db 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNMDGRNVNENVFINAPQSGTYYTIEVOAYN 420
QY 421 VPVGQPTFSLAIVN 434
Db 421 VPVGQPTFSLAIVN 434

RESULT 4
Q9AQR1 PRELIMINARY; PRT; 433 AA.
ID Q9AQR1 AC Q9AQR1
DT 01-JUN-2001 (T-REMBLrel. 17, Created)
DT 01-JUN-2001 (T-REMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-REMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. SD521.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133780;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SD521;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046405; BAB21268.1; -
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE66DDC CRC64;

Query Match 88.7%; Score 1993.5; DB 2; Length 433;
Best Local Similarity 87.6%; Pred. No. 1e-107;

Matches 380; Conservative 29; Mismatches 24; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
QY 61 NANTDNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSSGGLGSLPSNLQTLFSQAYS 120
Db 61 NANTDNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSSGGLGSLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAFGPTFILSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAFGPTFILSARSLAPDSSF 240
QY 241 WANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAQVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAQVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNMDGRNVNENVFINAPQSGTYYTIEVOAYN 420
Db 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNMDGRNVNENVFINAPQSGTYYTIEVOAYN 420
QY 421 VPVGQPTFSLAIVN 434
Db 421 VPVGQPTFSLAIVN 434

RESULT 5
Q9AQR4 PRELIMINARY; PRT; 433 AA.
ID Q9AQR4 AC Q9AQR4
DT 01-JUN-2001 (T-REMBLrel. 17, Created)
DT 01-JUN-2001 (T-REMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-REMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROA.
OS Bacillus sp. D6.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=127889;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=D6;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046402; BAB21265.1; -
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433

SQ SEQUENCE 433 AA; 45636 MW; 52087EOA2516107F CRC64;
 Query Match 88.5%; Score 1989.5; DB 2; Length 433;
 Best Local Similarity 87.3%; Pred. No. 1.7e-107;
 Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NNDNTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFSQAYS 120
 DB 61 NNDPNNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFSQAWN 119
 QY 121 AGARIHTNSWGAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGTFILSARSSLPDSSP 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTFTILSARSSLPDSSP 239
 QY 241 WANHDSKYAYMGTSNATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 240 WANYNSKYAYMGTSNATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADVGLGY 299
 QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGPKLKISLWSDAPASTTA 360
 DB 300 PEGDQGWGRVTLDKSLNVAYNVNEATLTGQKATYSFQAGKPKLKISLWTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTOYVGNDFTSFYNDWNGRNNVNFVINAPOSQGTITIEVOAYN 420
 DB 360 SVTLVNDLNLVITAPNGQYVGNDFSYDNNWNGRNNVNFVINAPOSQGTITIEVOAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSGQRFSLAIVH 433

RESULT 6
 Q9AQR2 PRELIMINARY; PRT; 433 AA.
 AC Q9AQR2;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Protease (Fragment).
 GN PROC.
 OS Bacillus sp. Y.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 RX MEDLINE=20568675; PubMed=1118284;
 RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
 RA Horikoshi K.;
 RT "Novel oxidatively stable subtilisin-like serine proteases from
 RT alkaliphilic Bacillus sp.; enzymatic properties, sequences, and
 RT evolutionary relationships.";
 RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL; AB046404; BAB21267.1; -;
 DR HSSP; Q45670; 1DBI.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004269; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 433 433
 FT NON_TER 433 433
 SQ SEQUENCE 433 AA; 45587 MW; B61291A803C775AE CRC64;
 Query Match 88.2%; Score 1982.5; DB 2; Length 433;
 Best Local Similarity 87.1%; Pred. No. 4.4e-107;
 Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NNDNTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFSQAYS 120
 DB 61 NNDPNNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFSQAWN 119
 QY 121 AGARIHTNSWGAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGTFILSARSSLPDSSP 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTFTILSARSSLPDSSP 239
 QY 241 WANHDSKYAYMGTSNATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 240 WANYNSKYAYMGTSNATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADVGLGY 299
 QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGPKLKISLWSDAPASTTA 360
 DB 300 PEGDQGWGRVTLDKSLNVAYNVNEATLTGQKATYSFQAGKPKLKISLWTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTOYVGNDFTSFYNDWNGRNNVNFVINAPOSQGTITIEVOAYN 420
 DB 360 SVTLVNDLNLVITAPNGQYVGNDFSYDNNWNGRNNVNFVINAPOSQGTITIEVOAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSGQRFSLAIVH 433

RESULT 7
 Q8T9W1 PRELIMINARY; PRT; 1825 AA.
 AC Q8T9W1;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Serine protease/ABC transporter TagD.
 GN TAGD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Atiyah C., Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF466309; AAL74253.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004269; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC_TM_transp.

```
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR000209; Peptidase S8.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC trans; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00136; SUBTILASE_SER; 1.
KW ATP-binding; Protease; Transport.
SQ SEQUENCE 1825 AA; 202641 MW; E28160BC78613A3B CRC64;

Query Match      22.5%; Score 506.5; DB 5; Length 1825;
Best Local Similarity 27.7%; Pred. No. 9e-21;
Matches 163; Conservative 77; Mismatches 155; Indels 193; Gaps 22;

QY 19 LYGGQGVAVADTGLDTR---NDS-----SMHEAFRGKITALYALGRTNNANDTNGH 68
DB 327 LRKGQQLSADTGLDGHCHFFSDSNPNIPYNSYNLNRKXVTVYIGSL--HDNEDYVDGH 384
QY 69 GTHVAGSVLNG-----STNKGAPOANLVFQSI-MDSGGGLGGLPSNLQLTFSOAYS 120
DB 385 GTHVCGSAAGAPEDSSLAISFSLATDAKIAFDLASDPSNNEPVPEDYSQLYOFLYN 444
QY 121 AGARIHNSGCA-----AVNGAYTTDSRVDVYRKX-DMTILFAAGNEGNGGTISAPGT 175
DB 445 AGARVHGDGWSLSTIQGLASYSDDAGSIDDFLYTHPDFILRAAGNNEQYSLLS-QAT 503
QY 176 AKNAITVGATENLRSP-----
DB 504 AKNIVTGAETHTHESITTDALYSNFEIVAKSTLNSLQSPDKDKYCTYTTAQCCTEYST 563
QY 193 -----GSYAD-----NINHVAQFSRGPTKDGRIKPDVMAPTFILSA 230
DB 564 VKLSGCGCTSYIKNSYASIFSSQPELYNENNICFSFSKGPHTDGLKPDVAPQCVITSA 623
QY 231 RSSLA-----PDSFHWANDSKYAMGTSNATPIVAGNVAQLREH----- 271
DB 624 RSNQANTDQCGSLGNLTALLUSE-----SGTSMATPLATAANTILQYLVGDGYPT 676
QY 272 --FYNRGITPKPILKAALIAGAADIGLGY-----NGNQMGWV 310
DB 677 GSIVESNKLQPTGSLKALINNAQLNGTFPLSNTNPSNAVDFDFAGANFVQWGS 736
QY 311 TLDKSLNVAYNSS-----SLTSOKATYST----- 338
DB 737 RMSEWL---YVESGVKPKPSRWVIGELGCKDKKASKNKEYSLSTGQNVSYCTYKPS 793
QY 339 --ATAGKP-LKISLVWSDAPASTTASVTLVNDLNLVITAPNGTQ---YVGNDFTPYND 391
DB 794 GSNSGGIPRIVATLVWTDPSYSGAKLNLVNDLNTW---NTESEFIFYNSNGSGSYNG 850
QY 392 N-----WDGRNNVNF---INAPSGTYTIEVQAYVVPVGPQPTS 429
DB 851 TKGTTPLQDSINNVEGIIVTPIPTNKGEISFRFIAGTNIPIGPQNF 898

RESULT 8
Q9GTN7 PRELIMINARY; PRT; 1702 AA.
AC Q9GTN7;
DC Q9GTN7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE TAGA.
GN TAGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
CX NCBI_TaxID=44689;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA Good J.R., Cabral M., Kuspa A.;
RT "TagA, a putative serine protease/ABC transporter of Dictyostelium
RT that is expressed at the onset of development and is required for the
RT differentiation of a subpopulation of prespore cells.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF263455; AAG11416.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001140; ABC TM transport.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR00209; Peptidase S8.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC trans; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;

Query Match      19.7%; Score 442; DB 5; Length 1702;
Best Local Similarity 25.6%; Pred. No. 4.4e-17;
Matches 139; Conservative 100; Mismatches 164; Indels 140; Gaps 22;

QY 18 GYGGQGVAVADTGLDTR---NDSSMHEAFRGKITALYALGRTNNANDTNGHGHVAG 74
DB 300 GIKGGEIVGCADTGIDINCHFFYDTNPIGSTRKIIIS-YSSGNGDQDDEIDGHTVIG 358
QY 75 SVLNGNSTN-----KGMAPQANLVFQSI-MDSGGGLGGLPSNLQLTFSOAYSAGARIHT 127
DB 359 TIIGTTVDPSVESEFGGAPNSKVAFDVLQVSGNGL-SIQSNLTATVQSYDQNAKVC 417
QY 128 NSGAAVNGAYTTDSRNVDDVYRKX-DMTILFAAGNEGNGG-----TISAPGTAKAHT 181
DB 418 DAWNSNIGFFYTGVTGTEMIDRFQWDHPDPLVYVRSAGN-NVNFGPSIVTLSQESTSKNSLV 476
QY 182 VGATENLRPSGYSADNINH-----
DB 477 VGSSN--QFS-STYLLSIDYDWDPIYNSIRTSVCTQCSYIGTICSDVPQTTSVDIQT 533
QY 202 -----VAQFSRGPTKDGRIKPDVMAPTFILS 229
DB 534 QCCSNPILAKICCSPEIQOQYQTNSTVYSEFIPSLFSGVGTSDGRLKPDLLAPGPSPIIS 593
QY 230 ABSSLAPDSSFWANH-----DSKYAMGTSMATPIVAGNVAQLRE-----HFVK 275
DB 594 SR-SLGPSSSTI--NHCSPTISGATISALIAEGSSQAAVAATSAALVLRQYRDGIFING 650
QY 276 R-----GITPKPSLLKAALIAGAA---DIGLGYPNNGQWGRVTLDKSLNVAYNSS-- 325
DB 651 KYNSSVGFQPSASLVKATLINTASINVDTLEY---SQGPGNIQLSKLITTNATQTTSLD 707
QY 326 -----SLTSOKATYSTATAGKPLKISLVWSDAPASTTASVTLVNDLNLVITA- 374
DB 708 IPSSTIEKADPIINTGETNSYCFSLDSKADIDITLVWTDPPAGSPLSFTFLVNNLDLALLAF 767
QY 375 PNG--TOYVGNDFTSYNDN---WDGRNNVNFNINAPSGTYTIEVQAYVVPVGPQPTS 429
DB 768 VDGELSIYSGNSETIFKNTSQVIFDQNLNNEVIRIKDAPIGSDVKIFGTNIVPNQSY 827
QY 430 LAI 432
DB 828 VVI 830
```

```
RESULT 9
Q8UB09 PRELIMINARY; PRT; 654 AA.
ID Q8UB09;
AC Q8UB09;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alkaline serine protease.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5010265; AL81794.1;
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ
  18.2%; Score 409; DB 17; Length 654;
  Best Local Similarity 29.6%; Pred. No. 1e-15;
  Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;
  18 GLYGGQIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGH 71
  154 GYDGGGITTIGDITGID-----ASHFDLQGV-----IGWDFVNGRSYPYDDHGH 202
  72 VAGSVLNGSTN---KGMAPQANLVFQSIM---DSGGGLGGLPSNLQTLFSAQYSAGARI 125
  203 VASIAAGTGAASNGKYKGAPAGKLAGIKVLGADGSGSISTIIKGVAVDNKKYGIKV 262
  126 HTNSWA-----AVNGAYTDSRNVDYVRKNDMTILFAAGNEGPGGTISAP 173
  263 INLSGSSQSSDGTALSAQVAAWDA-----GLVVVAAGNSGNGKYTIQSP 310
  174 GTAKNATVGTENLRPSFGSYADINHNVAQPSRRGPTKDGRIKPDVMAQGTIFLSASS 233
  311 AAASKVITGA-----VDKYDVITSFSSRGPTADGRLKPEVAVPAGNWIITAAAS 359
  234 LAPDSSFVANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAALTA 291
  360 ---GTSMGQIPNDYYTAAPGTSMATPHVAGIAALLIQ-----AHPSTPKVKTALIE 409
  292 GA-----ADTGLGYPNGQGWRTLDKSLNVAYNSSLSQVA-----TYSFT 338
  410 TADIVKPEADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKSGSTHQFV 460
  339 ATAGKPLKLSLVNSDAPASTASVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNN 398
  461 ISGASFVATLYNDNAN-----SDLDLYLDPNGNQ-VDYSYTAAY-----G 501
  399 VENVFNAPQSGTYTIEVQAVNVVPGQPTFSLAIN 434
  502 FEKVGYNPTDGTWTKVWSYS---GSANYQVDWS 534
  380 YVGNDFTSYNDNDGRNNVENFVINAPOSQTYTIEVQAY 419
  Q8RBJ2
```

```
Q8RBJ2 PRELIMINARY; PRT; 561 AA.
ID Q8RBJ2;
AC Q8RBJ2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Subtilisin-like serine proteases.
GN APRE2 OR TTE0824.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
[1]
SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
  Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
  Tan H., Chen R., Wang J., Yu J., Yang H.;
  "A complete sequence of T. tengcongensis genome.";
  Genome Res. 12:689-700(2002).
RL EMBL; AE013049; AAM24081.1;
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ
  17.9%; Score 401.5; DB 16; Length 561;
  Best Local Similarity 31.4%; Pred. No. 2.3e-15;
  Matches 143; Conservative 63; Mismatches 135; Indels 119; Gaps 22;
  6 GIVKADVAQSSYGLYGQIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTNAN-- 63
  155 GITK---ARSDFGVTGKNTIATIDIGDNHVDLS-----GGKI-----IGKDFINK 201
  64 ---DTNGHGHVAGSVLNGSTN---KGMAPQANLVFQSIMDSGGG-----LGLPSN 110
  202 TTPYDDNGHGHVAVASIAAGTGAGNSFYKGVAPDALLVGIKVLGDANGSGSMSTVTAGIDWA 261
  111 LQ-----TLPSQAYSAGARIHTNSWGAANGAYTDSRNVDYVRKNDMTILF 158
  262 VQNKDVYGIKVINLSLGTSTSSDG---TDSLSLVN-----RAVD-----SGIVVV 305
  159 AAGNEGPGGTISAPGTAKNAITVGTENLRPSFGSYADINHNVAQPSRRGPTKDGRIK 218
  306 AAGNSGPAKTYTIGSPGAEKAITVAAMADV---GELGFLN---ASFSSRGPTADGRIK 358
  219 DVMAQGTIFLSASSLAPDSSFWANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGI 278
  359 DIAPAGYNIITAAK-----ANSVNGYVYTSYSGTSMATPFVAGTVALMLN---ANPNL 405
  279 TPKPSLLKAALIAAGADIGLGYPNGQGWRTLDKSLNVA-----YNES 324
  406 TPNDAA--KNIMSTAKSWGPPSKXNDYAGRLQGYEARVAGFRGNIDVPNHYI--S 461
  325 SSLSTSKQA--TYSFTAT-AGKPLKISLV---WSDAPASTASVTLVNDLNLVITAPNGTQ 379
  462 GYLPGRSYSDTWTFTNATNTSYPIATLIIIPDWAYNP-----DFDIYDPSGTL 511
  380 YVGNDFTSYNDNDGRNNVENFVINAPOSQTYTIEVQAY 419
```



```

Db 173 EPYDD-NHGHGTHCAGDAANGALSDGQYQGPAPDANLVGVKVLNKTGS-GSLSTVIEGI- 229
QY 116 SOAYSAGARIHNTSMGAAVNGAYTTDSRVVDYVR-----KNDMTILFAAGNEGPNGGT 169
Db 230 DMCIQNQSYKNINILSL-GSDATEPASGDPVNAVETAWDNGVMVCAAGNSGPGDKT 288
QY 170 ISAPGTAKNAITVGATENLRPFGSYADINHVAFPSRGPYKQRIKPDVWAPGTFILS 229
Db 289 VQSPGISPKVITVGAADDNNNTAERS-----DSDVAEFSSRGPITDGLTKPNLLTPGVDIVS 344
QY 230 ARS--SLAPDSFWANHDSKYAYMGGTSMATPIVAGNVQALREHFVKNRGITPKPSLLKA 287
Db 345 LRAPGSFIDKTNKSARVGSNYLSLGSNTAIPICIGIVAILQ---SDSLT--PNQVKE 399
QY 288 ALIAGAADTGLGPNQNGRVTLDKSLNVAYVNE 323
Db 400 KLMEACQDLQGS-PN-VQAGYLV---NAANLININE 430

RESULT 13
Q9FC06 PRELIMINARY; PRT; 1253 AA.
AC Q9FC06;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative secreted peptidase.
GN SC07176 OR SC8A11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Frazer A., Goble A., Hdaigo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RP "Complete genome sequence of the model actinomycete Streptomyces
RP coelicolor A3(2).";
RL Nature 417:141-147(2002).
RL EMBL; AL939130; CAC01576.1; -.
DR HSSP; Q99405; IMPT.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```

```

DR InterPro; IPR001317; PA.
DR InterPro; IPR000209; Peptidase_s8.
DR Pfam; PF02225; PA; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILISIN.
DR PROSITE; PS00137; SUBTILISIN.
DR PROSITE; PS00138; SUBTILISIN.
KW Complete proteome.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EFEEB89 CRC64;

Query Match
Best Local Similarity 27.4%; Score 371; DB 16; Length 1253;
Matches 143; Conservative 59; Mismatches 145; Indels 170; Gaps 22;

QY 16 SYGLYGGQIVAVADTGLDTRNDNSMHEAFRKITALLYALGRTNNTDNGHGHVAGS 75
Db 233 SGGNTGEGVGAVLDTGVDAG-----HFDFAGRIAATASFVPDQDVTDRNGHGHVAST 286
QY 76 VLNGNSTN---KGMAPQANLVFQSIMDSGG
Db 287 VAGTGAASGVGEKGVAFGLSHLGHKVKVLDNSGQSDSWVLAGWEAVRDQHAKIVNSLGD 346
QY 107 LPSNLQTLFSQAYSAGARIHNTSMGAAVNGAYTTDSRVVDYVRKNDMTILFAAGNEGPN 166
Db 347 SPTDGTDLPLSEAV-----NWLGAETGA-----LFVVAAGNSGPE 380
QY 167 GGTISAPGTAKNAITVGATENLRPFGSYADINHVAFPSRGP-TKDGRIKPDVWAPGT 225
Db 381 AYTGTGPAADAALTGVAVNG--PGKG-----VDQLADFSSRGPRVGDNAVRDLTPAGV 433
QY 226 FILSARSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVQALREHFVKNRGITPKPSLL 285
Db 434 GVLAARSRYAPEG-----EGAYQSLSGTSMATPHVAGAAALLAAEHPDWTG-----QRL 482
QY 286 KALLIAGAADIGLYPENGNO-----GNGRVTLDKSLNVAYVNESSLSLTSQ---KATY 335
Db 483 KEALVGTTA-----GTQRFSPFDAGSGRV-----DVAAAVRSLTLAGSDAFQAHY 528
QY 336 -----SFTATAGKPLKISLVWSDA-----PASTTASVTLVND 367
Db 529 PYTPGQTVRRDVTYNTSGPAPFVALDLALSPAELPEGLFTLSBAQVTVPAHGTSVGVIH 588
QY 368 LN-----LVITAPNGTQYVNDFTSPYNDWDRNVDGNVNFVNA-----PQSG 410
Db 589 LDAEDNGAYATLVASGADGAVLA----RTPVGVNKEGR--ATLALTAKDHHKPLSG 642
QY 411 T-----YTIEVQA-YNVVGVGQTFSL 430
Db 543 TVILKQVETNATPKVYSDASGELDRLSPSTYSV 677

RESULT 14
P95684 PRELIMINARY; PRT; 1102 AA.
AC P95684;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Subtilisin-like protease.
OS Streptomyces albobacillus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-3253;
RX MEDLINE=97144528; PubMed=8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
RT Streptomyces albobacillus.";
RL J. Bacteriol. 179:430-438(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 19:18:46 ; Search time 3605 Seconds
(without alignments)
5217.996 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVOAYNPVGPQTFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/csm2_1/USPTO_epool/SWOP985689/runat_10032004_112806_19360/app_query.fasta_1.583
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=SWOP985689.ecgn_1_1_2372 @runat_10032004_112806_19360 -ICPU=3
-NO_MAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2242	99.8	1305	6	AX839476	Sequence
2	2242	99.8	1923	1	AB051423	Bacillus
3	2242	99.8	1923	6	AB368117	Sequence
4	2237	99.6	1923	1	AB084155	Bacillus
5	2237	99.6	1923	6	AB368118	Sequence
6	2178	96.9	1920	1	AB046403	Bacillus
7	2178	96.9	1920	6	AB368116	Sequence
8	2138	95.1	1302	1	AB046406	Bacillus
9	2120.5	94.4	3003	6	AR069954	Sequence
10	2120.5	94.4	3003	6	BD062155	Nucleic a
11	1993.5	88.7	1299	1	AB046405	Bacillus
12	1989.5	88.5	1299	1	AB046402	Bacillus
13	1982.5	88.2	1299	1	AB046404	Bacillus
14	1981.5	88.2	2218	6	EO3808	EO3808 DNA encodin
C 15	627.5	27.9	60006	1	AF268611	Uncultured
16	585	26.0	301205	1	AF017218	Geobacter
17	519.5	23.1	5890	3	DDU60086	Dictyosteli
18	506.5	22.5	5772	3	AF466309	Dictyosteli
19	492	21.9	6115	3	DDU20432	Dictyosteli
20	447.5	19.9	1977	6	AR201152	Sequence
21	442	19.7	5109	3	AF263455	Dictyosteli
C 22	420.5	18.7	126928	2	AC096673	Trypanoso
23	411.5	18.3	1236	6	AR201146	Sequence
24	411.5	18.3	1962	6	AR201155	Sequence
C 25	411.5	18.3	12452	1	AE010265	Pyrococcu
C 26	401.5	17.9	14376	1	AE013049	Thermoana
C 27	399	17.8	303450	1	SC0939130	Streptomy
C 28	381	17.0	29850	1	AF004601	Oceanobac
29	366	16.3	3348	1	D83672	Streptomyce
30	363.5	16.2	300425	1	AP005044	Sequence
31	353	15.7	2539	6	AR202322	Sequence
32	349	15.5	291000	1	SC0939105	Streptomy
33	348	15.5	2809	6	AR202321	Sequence
34	348	15.5	3900	1	AB007809	Streptomy
C 35	348	15.5	302300	1	AF005034	Streptomy
36	347.5	15.5	292550	1	AF001513	Bacillus
37	338	15.0	1239	1	AF305633	Thermoana
38	338	15.0	1239	1	AY028704	Thermoana
C 39	337	15.0	12295	1	AE013026	Thermoana
40	336	15.0	135638	1	AF484556	Streptomy
C 41	332.5	14.8	1329	6	AX433519	Sequence
C 42	331	14.7	207829	1	BSUB0010	Bacillus su
C 43	329.5	14.7	6854	1	SCU39230	Bacillus sp
C 44	329.5	14.7	300800	1	SC0939112	Streptomy
C 45	327.5	14.6	301150	1	AP004602	Oceanobac

ALIGNMENTS

RESULT 1

AX839476
LOCUS AX839476 1305 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 Okuda, M.K., Saito, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
Alkaline protease
Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
Location/Qualifiers
1..1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
<1..1305
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAE85146.1"
/db_xref="GI:39922767"
/translation="NDVARGIVKADYAGSSVGLYQGGIIVAVADTGLDTRGNDSSMHE
APRKITALYALGRINNDINGHTHVGSLNGSTNGKNGAPQANLVFOSIMDSGG
GUGLPSNLQIFSGAGARHTNSWGAAYTTDSRVDDYVRKNDMTILFAA
GNEPNGTISAPGAKNAITVGATENRPSGYADNINHWAFSSRPTKDGRIKP
DYVAPGTIFLSASLSLADPSFWANHSKYVMGCTSNATIVAGNVAQLRHFVYKR
GITPKPSLLKAALIGAADIGLYPNQNGWGRVTLDKSLNVAVYVNSSLSLTSOKAT
YSFTAGKPLKISLVMSDAPASTTASVLTVDLDELVTAPNGIYVGNDFTSFYNDN
WGRNVNENFINAQSGTYTIEVQAYNVVPGPQFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,19e-133 Length: 1305
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.78% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1-COPY (1-434) x AX839476 (1-1305)
QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATTTGGCGTGGGAATTTCTCAAGCGGATGTGGCTCAGAGCAGCTACGGTTGTAT 60
QY 21 GlyGlnGlyLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGACAAAGGACAGATCGTAGCGTTCGCCGATACAGGCTTGATACAGGTTCGAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCGATCATGAGCCCTTCGCGGGAAAATTTCTGATATATATGCTATATGCTGGCGGACGAAT 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCCAATGATACGAATGCTCATGTGTACGATGTGGCTGTGGCTCCGTATTAGGAAACGGC 240
QY 81 SerThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 241 TCCACTTAATAAGGAATCGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGTGATAGC 300
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 301 GGTGGGGGACTTGAGGACTACCTTCGATCTGCAACCTTATTTCAGCCAGCATACAGT 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
DB 361 GCTGGTGCCAGATTTCATCAAACTCTCGGGGAGCAGCAGTGAATGGGGCTTTACAAACA 420

QY 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
DB 421 GATTCACGAAATGCTGATGACTATGTGCGCAAAAATGATATGACGATCTCTTTTCGTGCC 480
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
DB 481 GGGAAATGAAGGACCGAACCGGAAACCATCAGTGCACCGGACACAGCTAAAATGCAATA 540
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 541 ACAGTCGGAGCTACGGAAAACCTCCGCCCAAGCTTTGGCTTATCGGACATATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleIysProAspVal 220
DB 601 CATGTGGCACAGTTCTCTTACGTGACCGCAACAGATGGCGGATCAACCGGATGTC 660
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 661 ATGGCACCGGAAAGCTTCACTATCAGCAAGATCTTCTTGGACCGGATCTCTCTCTTC 720
QY 241 TrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 721 TGGCGCAACCATCACAGTAAATATGATACATGGTGGAAACGCTCCATGGCTACACGATC 780
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
DB 781 GTTGTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAACAGAGGCATCACACCA 840
QY 281 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 841 AAGCTTCTCTATTAATAAGCGGCACCTGATTCGGTGGCAGTCACATCGGCTTGGCTAC 900
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr 320
DB 901 CCGAACGTTAACCAAGATGGGACGAGTGACATTCGATTAATCTCCTGAACGTTGCTAT 960
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 961 GTGAACGAGTCCAGTTCTCTATCCACGCAAAAGACGACGCTACTCTGTTTACTGTACT 1020
QY 341 AlaGlyIysProLeuIysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
DB 1021 GCCGGCAAGCCTTTGAAAATCTCCCTGGTATGCTGATGGCCCTCGAGCACAACCTGT 1080
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1081 TCCGTAAACGCTTCTCAATGATCTGGACCTTGCTATTACCGCTCCAATGGCACAGTAT 1140
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1141 GTAGGAAATGACTTTACTTCGCCATCAATGATTAATCTGGGATGGCGCAATACGTAGAA 1200
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1201 AATGTTATTATTATGACACCAAGCGGAGCGTATACAAATTGAGGTACAGGCTTATAAC 1260
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1261 GTACCGGTTGGACACAGACTTCTCGTTGGCAATTGTGAAT 1302
RESULT 2
AB051423
LOCUS AB051423 1923 bp DNA linear BCT 10-MAY-2002
DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
ACCESSION AB051423
VERSION AB051423.2 GI:20521154
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 Itoh, S. and Saeki, K.
AUTHORS

TITLE new protease
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1923)
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606, AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@kastonet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKVFSLVLSAAILSTVALSNPSAGGARNFPLDPKQIQT
 TDKGFQKQGTGAAPLSENVKLPKGLQKLETPANNKHLIIQFNGPILSEETKQ
 QLEKTAKILDIYIPDYAVIYVEYEDGVKSTSTIEHVESPEYPIYRIDPQLFTKAS
 ELVKAVALDTKQKNKEVQLRGIEQIAQFAISNDVLYITAKPEYKVNVDVARGIVKADV
 AQSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNVANDTNGHGT
 HVAAGLVNGSTNKGMAPOANLVQSIMDGGGLGPNLQTLFSCAYSAGARIHTN
 SWGAANVGATTDNRNVDVVRKNDMTILFAAGNEGPNGSTISAPCTAKNAITVGATE
 NLPSFGSYADNINHVAFSSRGPTKDGRIKPYMAPGPTFLSARSLAPDSFVANH
 DSKYANGSTMAIPVAGNVQIREHFVKNRGIPTKPSLLKKAALTAGAADIIGLGP
 QNGWGRVTDLSKNAYVNESLSSTSKATYSFTATAGKPLKILSWSDAPATPA
 SVTLNDLDLIVITAPNGTVYQVNDFTSPYNDNWDNRNVENFINAPQSGTVEVQA
 YNVPVGPQTFSLAIVN"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.92e-133 Length: 1923
 Score: 2242.00 Matches: 433
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.78% Indels: 0
 DB: 1 Gaps: 0
 US-09-985-689A-1-COPY (1-434) x AB051423 (1-1923)
 Qy 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTCGCGTGAATTGCAAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnLeuAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAGGACAGATCGTAGCGGTGGCGATACAGGGCTTGATACAGGTGCGAATGACAGT 738
 Qy 41 SerMethisGluAlaPheArgGlyLeuIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCTTCCCGGGAAATTTACTGCAATTATGCAITGGACGACGAAT 798
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTCATGTGTACGATGTGGCTGGCTCCGTATTAGAAACGGC 858
 Qy 81 SerThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGAAATGGCGCTCAGCGCAACTAGTCTCCAACTATCATGGATGAGC 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGAGGACTACCTTCGAATCTGCAACCTTATTAGCAGCAAGCATACAGT 978

Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 979 GCTGCTGCCGAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCCAGAAATGTGGATGACTATGTGGCAAAATATATAGCAGATCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 Db 1099 GGGATGAAGGACCGACCGGGAACCATCAGTCAGCCAGGACAGCTAAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAAGCTTTGGGTCTTATGCGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGATCTCTTCACGTGGACCGAAGGATGACCGGATCAACCCGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerIleAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCAGGAACTTCATCTACTATCAGAAAGATCTTCTCTGACCGGATTCCTCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGAAACCATGACAGTAAATATGCATACATGGGTGGAACGTCCATGGCTACCCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
 Db 1399 GTTGTGGAACCTGGCACAGCTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 Qy 281 LysProSerLeuLeuLeuAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAGCGGCACTGATTCGGGTGAGTGCAGTGCAGCTGGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGCTAACCAAGGATGGGACGAGTGAGTGGATTAATCCCTGAACGTTGCCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACGACCAAAAGCGAGCTACTCGTTTACTGTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGATGGTCTGATGCCCCCTGGAGACACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATTAACCTGGGATGGCCGCAATACGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTATTTATTAATGCACCAAAAGCGGACGTATACATTAAGTACAGGTATATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGTGAAT 1920
 RESULT 3
 LOCUS AR368117 1923 bp DNA linear PAT 12-SEP-2003
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.

Unclassified.
1 (bases 1 to 1923)
Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kegeyama,Y.,
Shikata,S. and Nomura,M.
Alkaline protease
Patent: US 6376227-A 5 23-APR-2002;
Location/Qualifiers
1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,92e-133 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 AATGATGTTCCGGTGGAAATTTCAAGCGGATGGCTCAGACAGCTACGGGTGTAT 678
Qy 21 GlyGlnGlyLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGGACAGATCGTAGCGGTTGCCGATACAGGCTTGATACAGGTCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCATGAAGCCCTCCCGCGGAAATTAATCTGCATTAATATGCAATGGGACGGAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGCTCATGTACGATGTGGCTGGCTCGGTATAGGAAACGGC 858
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCATCTATCATGTATAGC 918
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 919 GTGTGGGGACTTGGAGGACTACCTTCGATCTGCAACCTTATTCAGCCCAACATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 979 GCTGGTGCCAGAAATTCATAAACTCTGGGGACGACGAGTGAATGGGCTTACACAACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaIle 160
Db 1039 GATTCAGAAATGTGATGACTATGTGGCNAATAATGATATACGATCCTTTTCGCTGCC 1098
Qy 161 GlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGGAAATGAAGACCGGAACGGCGAACCATCAGTCGACACGACAGCTAAAAAATGCATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTGGAGCTACGGAACCTTCCGCCCAAGCTTTGGGTCTTATGCGGACAAATATCAAC 1218
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAAAGGATGGACGATCAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1279 ATGGCACCGGAACGTTCTACTATCATCAGCAAGATCTTCTTTCGACCGGATTCCTCCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyIleThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAAATATGATACATACATGGGTGGAACTGCTCATGCTACACCGATC 1398

261 ValAlaGlyAsnValAlaGlnLeuArgGlyHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTGTGGAAACGTGGCACAGCTTCTGTAGCATTTTGTGAAACACAGAGCATCACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAAGCGGCACATGATTGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACGGTAACCAAGATGGGACAGTGACATTCGATAAATCCCTGACCGTGGCTAT 1578
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1579 GTGACGAGTCCAGTTCCTCTATCCACGACCAAAAGCGACGACTCTGTTACTGCTACT 1638
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGAGACCACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1699 TCCGTAACGTTGTCAATGATCTGGACCTTGTCATTACCGCTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1759 GTAGGAATACATTTACTTCGCATCAATGATACTGGGATGGCGCAATAACGTAAGA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGATTTATTAATGCCACCAAGCGGACGCTATACATTTGAGGTACAGGCTTATAAC 1878
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACACAGACCTTCTCGTTGCCAATTGTGAAT 1920

RESULT 4
AB084155
LOCUS AB084155
DEFINITION Bacillus sp. KSM-9865 gene for protease, complete cds.
ACCESSION AB084155
VERSION AB084155.1 GI:34392386
KEYWORDS
SOURCE Bacillus sp. KSM-9865
ORGANISM Bacillus sp. KSM-9865
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
1
Okuda,M., Saeki,K. and Kobayashi,T.
Bacillus sp. KSM-9865 protease gene
Published Only in Database (2003)
2
(bases 1 to 1923)
Okuda,M., Saeki,K. and Kobayashi,T.
Direct Submission
Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,
Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan E-mail:okuda.mitsuyoshi@kao.co.jp.
Tel:81-285-68-7543, Fax:81-285-68-7547)

FEATURES
Location/Qualifiers
1..1923
/organism="Bacillus sp. KSM-9865"
/mol_type="genomic DNA"
/strain="KSM-9865"
/db_xref="taxon:192495"
1..1923
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAC82522.1"
/db_xref="GI:34392387"
/translation="MRKKKVLFLVLSAAAILSTVALSNPSAGARNFOLDPKGIOTI
TDKGFSGKQQTGAATAAFVSESNKLPKGLQKLETPANNKLHIQVNGPILEETKQ
QLEKTKAKILDYIPDVAYIVEBGDKVKSATSTIEDVESVEPYLPVIRIDPQLFTKGAS

ELVXAVALDTNKKKEVQLRGIEQIAQFATSDVLYITAKPEYKVMNDVARGIVKADV
 AQSSTGLYGGQGI VAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNANDTNGHT
 HVASVLGNGTNGMGAPOANLFGSDMGGGLGGLFSLNQLTFLSOAYAGAGARIHTN
 SWRAVNGAYTTDSRNVDYVYRKDMTILFAAGNEGPNGGTISAPGTAKNAITVGATE
 NLRSPFSYADINIHVAQFSRGPTKDGRIKPDVMAFGPTFLSARSLAPDSFVWANH
 DSKVAYMGTSMAPIVAVAGNVAQLREHFVKNRGITPKPSLLKALJAGADIGLVEN
 GNOGMRVTLKSLNVAVNSSLSLSTOKATYSPTATAGKPLKISLWSDAPASITA
 SVTLNDLIDLITAPNGTQYVGNDFTSFYNNWMDGRNVNFNAPQSSTGYTIEVQA
 YNVPVGPPTFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 3,99e-133 Length: 1923
 Score: 2237.00 Matches: 432
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 99.54% Mismatches: 0
 Query Match: 99.55% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGCACGTGGAAATTTCAACCGCATGTGGCGACAGCAGCTACGGGTTCTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACACGACAGATGTCAGCGTTGCCATACAGGGCTTGATACAGTTCGAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAAGCCTTCCCGGGGAAATTTACTGCATTATATGCAITGGACGCGCAAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCATGATACGAATGGTCATGGTACGATGTGGCTGGCTCCGTATTAGGAAACGGC 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTATAAGGAATCGCGCTCAGCGCAATCTAGTCTTCCAAATCTATCATGATGATG 918
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGAACTCGAACTTATTCAGCAGCATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 979 GCTGGTGCAGAAATTCATAAACTCTCGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGATGATGTGCGCAAAATGATATGACGATCCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GCGAATGAAGCAGCAGCGCGAACCATCATAGTCACCCAGCAGCATAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGAGACTACGAAACCTCGGCCAAGCTTCGGGTCTTATGGCGCAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCACAGTCTCTTCAGTGGACCGCAAGGNTGGACGATCAACACCGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGACCGGAAACGCTTACTATCATCAGCAAGATCTTCTCTTGCACCGGATCTCTCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGGCAACCATGACAGTAATATGATCATATGGTGGACGTCATCGGCTACACCGATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTATTAAAGCGCACTGATGTCGGTGGCAGCAGACATCGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCTTAACGTTGCCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGCTCAGTCTCTATCCACCACCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCGCGACAGCCTTTGAAGATCTCCCTGGTATGCTGTGATGCCCTCGCAGCAACAACCTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCGTAAACGCTTGTCAATGACCTGGACCTTGTCAATACCGCTCCAAAATGGCACACATAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTTGGAAATGACTTTACTTTCGCATACAAATAATACTGGGATGGCGCAATAACCGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTAATTTAATATGCGCCACAAAGCGGAGCGTATACAAATTGAGGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATTGTGAAC 1920

RESULT 5

AR368118 1923 bp DNA linear PAT 12-SEP-2003
 LOCUS Sequence 7 from patent US 6376227.
 DEFINITION AR368118
 ACCESSION AR368118
 VERSION AR368118.1 GI:34601779
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1923)
 AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
 Shikata, S. and Nomura, M.
 TITLE Alkaline protease
 JOURNAL Patent: US 6376227-A 7 23-APR-2002;
 FEATURES Location/Qualifiers
 source 1..1923
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Alignment Scores: 3,99e-133 Length: 1923
 Pred. No.: 2237.00 Matches: 432
 Score: 2237.00
 Percent Similarity: 99.77% Conservative: 1
 Best Local Similarity: 99.54% Mismatches: 1
 Query Match: 99.55% Indels: 0
 DB: 6 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AR368118 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATTTCAACCGGATGGCTCAGAGCAGCTACGGTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGCACAGGACAGATCGTAGCGGTTCGCCGATACAGGCGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAAGCCCTCCGCGGGAATTAAGTGCATTATATGCAATGGAGCGAGCAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGGTCTCATGGTAGCGATGGCTGGCTCGCTATTAGGAAACGGC 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATGGCGCTCAGGGCAATCAGTCTTCAATCATATGATGATAGC 918
 QY 101 GlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGGAGGACTTACCTCGAATCTGCAACCTTATTGAGCCCAAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValAlaTyrThr 140
 Db 979 GCTGGTGCCGAATTCATCAAACTCTGGGAGCGCAGTGAATGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGCGATGACTATGTGCGCAAAATGATGACGATCCTTTTCGCTGCC 1098
 QY 161 GlyAsnGlyLeuProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGAATGAAGGACCGGACCGGCAACCATCAGTCACAGGACAGCTGAAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGATTCTTTCAGGTGACCGGACGACAAAGATGGACGATCAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGGAAGCTTACATCTATCATCAGCAAGATCTTCTTCGACCGGATTCCTCCTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAAATATGATATCATGCGGTGGAACTCCATGCTACACCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAATAACAGAGGATCACACA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATGCGCGGTGACGCTGACATCGCCCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgGlyThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGACGGTAACACAGATGGGACGAGTGACATTGATGAATAATCCTGACGCTTCCCTAT 1578
 QY 321 ValAsnGlySerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACAGCAAAAGACGACGTACTCGTTTACTGCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCGCCCTGCGAGCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCGGTAAAGCTTCAATGATCTGGACCTTGTTCATTACCGCTCCAAATGGCACACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1759 GTAGGAATGACTTTTACTTCGCGCATACATGATTAAGTGGATGGCGGCAATAACGTAGAA 1818

QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATATTATTATGACCAACAAAGCGGACGTATACATTTGAAGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920
 RESULT 6
 LOCUS AB046403 1920 bp DNA linear BCT 10-MAY-2002
 DEFINITION Bacillus sp. 9860 PROA gene for protease, complete cds.
 ACCESSION AB046403
 VERSION AB046403.2 GI:20521152
 SOURCE
 ORGANISM Bacillus sp. 9860
 Bacillus sp. 9860
 Bacteria, Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 AUTHORS Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 1118284
 REFERENCE 2 (bases 1 to 1920)
 AUTHORS Saeki, K.
 DIRECT SUBMISSION
 SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Hage, Tochigi 321-3497, Japan (E-mail: 387185@katsanet.kao.co.jp, Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 COMMENT On May 9, 2002 this sequence version replaced gi:12381938.
 FEATURES
 source
 1..1920
 /organism="Bacillus sp. 9860"
 /mol_type="genomic DNA"
 /strain="9860"
 /db_xref="taxon:133778"
 gene
 1..1920
 /gene="PROA"
 CDS
 1..1920
 /genes="PROA"
 /protein_id="BAB21266.2"
 /product="protease"
 /transl_table=11
 /codon_start=1
 /protein_id="BAB21266.2"
 /db_xref="GI:20521153"
 /translation="MRKKVFLSVLSAAAILSTVALNPSAGDARTFDLDFKGIQTTT
 DVSGFKRQTGAALFVSENVKLLKLETPANNKLHIVQNGPILEETKQK
 LETTGAKILDYIPDAVIVVEGDVQSVKSTIEHVSVPYLPKYKIDPOLPTKGAST
 LVKALADTKQNNKVLQRIEIAQYASNDVHYITAKPEYKVNNDVARGIVKADVA
 QSYGLYGGQQLVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTNANDTNGHTS
 VAGSVJNGKATNGMAPQANLVFQSIQSSGGGLGSPNLQTLFQSFAGARIHTNS
 WGAANGAYTDSRNVDDYVRKNDMTILFAAGNERNGGTISAPGTAKNAITVGATEN
 LRPSGYSADININHAQFSRGTQDRIKPDVMAFGTYILSARSLAPDSFWANHD
 SKYVMGTMATPIVAGNVAQLRHFVKRGITPKPSLLKALIAAADVGLGYENG
 NOGWGVTDKSLNAVYNESALSTSQKATVTFATAGKPLKISLWSDAPASTTAS
 VTLVNDLDVITAPNCTRYVGNDFSAFPDNNNGRNVENVINSQSQSYTILEVQAY
 NVFVGPQNPSLAIV"
 ORIGIN

Alignment Scores:
 Pred. No.: 2,22e-129 Length: 1920
 Score: 2178.00 Matches: 417
 Percent Similarity: 99.31% Conservative: 14
 Best Local Similarity: 96.08% Mismatches: 3
 Query Match: 96.93% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AB045403 (1-1920)

```
Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGATATTGTCAAAGCGGATGTGGCAGAGCAGTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGGCCAGATTGTCCGAGTTGGGATCTGGATTGGATACAGAAAGACGACGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrzAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCAATGAAGCTTCCCGGTAAATAAACAGCACATATATGCACTGGGTGGACCAAT 795
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGATGATGATACCATCTGGCAGTTGGTATTAGAAATGCG 855
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAAGCAATGGCACCTCAAGCGAATCTGTTTTTCAATCCATCATGGATAGC 915
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGTGGAGCTTGCCTTCCATCTCAAAACCTTATTCAGCCAAAGCATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThr 140
Db 976 GCAGGTGCCAGATTATCATCAAACTCTGGGGGCGCGGTGAATGGGGCTTACAGCACA 1035
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GATTCCAGAAATGGGATGACTATGAAGAAAAATGATATGACGATTCTTTTCGGGGT 1095
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGATGAAAGCGCGAAGCGCGTACCATCGTGCACCTGGTACGCTAAAGCGCCATA 1155
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCAACCGGAAACCTCGCGTCCAGCTTGGTTCTTATGCAGATAATATAAC 1215
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCAGATTCTCTTCCCGTGGCCGACAAAGATGGGCGAATCAAGCCCTGATGTC 1275
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCAGGACATACATATTTATCAGCAAGATCTTCTTTCACCCGATTCCTCCTTC 1335
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACAGCAATATGCTATATGGTGGAAACGTCCTCATGGCAACCGCAT 1395
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTGCACAGCTCCGTGACATTTTGTGAATAATAGAGGAATCACTCCT 1455
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCTTTCCCTATTGAAAGCAGCTTTGATTGCGAGGTGCTCTGATGTGGATTGGGTTAT 1515
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGGAACCAAGGATGGGCGGAGTAGCTCCCTGGATTAATCGTTGAACGTGGCTAT 1575
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACCAATCCAGTGCCTTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
Db 1636 CGGGCAAGCCATGGAATAATCTCCTGGTATGTGCGATGCCCTCGAAGCACTACTGCT 1695
```

```
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTTGGTCAATGATTTGGATTGGTCAATTACAGACCAACGGAAGATAT 1755
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
Db 1756 GTCGGGAATGACTTCTCAGCACCATTTGACAATACTGGATGGCCGCAATAACGTAGAA 1815
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTATTATTAATTCGCCCAAGTGGAAACATATACCATTTAGGTGCAAGCATATAAT 1875
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACCACAAAACCTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 7
AR368116 1920 bp DNA linear PAT 12-SEP-2003
SEQUENCE 3 from patent US 6376227.
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S., and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,22e-129 Length: 1920
Score: 2178.00 Matches: 417
Percent Similarity: 99.31% Conservative: 14
Best Local Similarity: 96.08% Mismatches: 3
Query Match: 96.93% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1-COPY (1-434) x AR368116 (1-1920)
Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGATATTGTCAAAGCGGATGTGGCAGAGCAGTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGGCCAGATTGTCCGAGTTGGGATCTGGATTGGATACAGAAAGACGACGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCAATGAAGCTTCCCGGTAAATAAACAGCACATATATGCACTGGGTGGACCAAT 795
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGATGATGATACCATCTGGCAGTTGGTATTAGAAATGCG 855
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAAGCAATGGCACCTCAAGCGAATCTGTTTTTCAATCCATCATGGATAGC 915
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGTGGAGCTTGCCTTCCATCTCAAAACCTTATTCAGCCAAAGCATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThr 140
```

```

Db 976 GCAGGTGCCAGATTTCATACAACTCTCTGGGGGAGCGGTGAATGGGGCTTACAGACA 1035
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleuPheAlaLa 160
Db 1036 GATTCAGAAATGTGATGACATATGAAGAAATATGATGACCATCTTTTCGGGCT 1095
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLe 180
Db 1096 GGGAAATGAAAGCGGACCGGTACCATCAGTCACCTGGTACCGCTAAACACCCATA 1155
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCACACCAACCTCGCTCCAAAGCTTCGGTTCCTATGTCAGATTAATTAAC 1215
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CAGCTTGACAGTCTCTCTCCGTGGCCGACAAAAGATGGCGGAATCAAGCTCATGTC 1275
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGGACATACATATTTATCAGCAAGATCTCTTGCACCCGATTCCTCCTTC 1335
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACAGCAATATGCTATATGGGTGGACGTCCTCAATGCAACACCGAT 1395
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCTTCTCCATATGAAAGCAGCTTTGATTCAGGTGCTGCTGATTTGATGGGTAT 1515
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGAAACCAAGATGGGCGGAGTGACCTGGATAAATCGTGAACGTGGCTAT 1575
Qy 321 ValAsnGlySerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGACCAATCCAGTGGCTTATCACTAGCCAAAGGACATATACCTTTACTGCAACG 1635
Qy 341 AlaGlyLysProLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCATTTGAAATCTCCCTGGTATGTCGGATGCCCTCAGACACTACTGCT 1695
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTGGTCAATGATTTGGATTTGGTCAATACAGCACCAACGAAACAGATAT 1755
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1756 GTCGGGAATGCTTCTCAGACCATTTGACAAATACCTGGATGCCGCAATACGTAGAA 1815
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTTATTAATTCGCCCAAGTGGAAACATATATACCATTTGAGTGGCAAGCATATA 1875
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACCAACAACTTCTCGTTGGCAATTTGTGAC 1917

```

RESULT 8

```

AB046406
LOCUS AB046406 1302 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
ACCESSION AB046406
VERSION AB046406.1 GI:12381944
KEYWORDS Bacillus sp. NV1
SOURCE Bacillus sp. NV1
ORGANISM Bacillus sp. NV1
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

REFERENCE 1 (sites)
AUTHORS Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1302)
AUTHORS Saeki, K.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:1387185@kastanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
Location/Qualifiers
1..1302
/organism="Bacillus sp. NV1"
/mol_type="genomic DNA"
/strain="NV1"
/db_xref="taxon:133781"
1..1302
/gene="PROE"
<1..>1302
/gene="PROE"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21269.1"
/db_xref="GI:12381945"
/translations="NDVARGIVKADVQAQSSYGLYGQGVVAVADTGLDTRGNDSSMHE
AFSGKTAIVAGRTNNDPNHGTHVAGSVLNGTSNKGMAPOANLVFQSVYVDSNG
GLGGLPSNVSTLSQAYSGARITHNSWGPVNGAYTDSRVNDDYVRKNDMAVLFAA
GNGPNGGTATSAFCTAKNALTGTATENLRFSGSYADNINHVAQFSRGTCKGRKXP
DNVAPGFILUSASLAPDSSFWANHDSKYAYMGTSMATPIVAGNVAQKRFIKNR
GITPKSLLKALTAGATDGLGYPGNGQGRVTLDKSLNVAFVNETSLSTNQKAT
YSFTAQSGKPLKISLVMSDAPASTASVTLVNDLVLITAPNGTKYGVNDFTPAYDNN
WDGRNNVENVPINAPSGGTYTEVQAYNVNPPQGFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 4,75e-127 Length: 1302
Score: 2138.00 Matches: 405
Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 9
Query Match: 95.15% Indels: 0
DB: 1 Gaps: 0
US-09-985-689A-1-COPY (1-434) x AB046406 (1-1302)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSertyGlyLeuTyr 20
Db 1 AACCATGTTGCCAGAGCATTTGTAAGCCGATGTTGCCAGAGCAGCTATGTTTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGGCAAGGCGCAAGTGGTTCAGTAGCAGTACCGGACTGGATACAGGCGGTAAATGACAGC 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGTCATGAAGCGTTCCGAGGAAAGATTACCGCATATATGCCCTTGGAGAACAAAC 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AACCCCAATGATCAAAACGGACACGGACCGCATGTTCCCGATCTGTTTAGGAAACCGT 240
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 241 ACTTCGAATAAAGGATGGCTCCACAGCTAACCTAGTTTCCATCTGTTATGACAGC 300
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

```

```
Db 301 AATGGCGGACTTGGCGGACTGCTTCCATGTAAGTACATATTATTCAGCCAGGCATATAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 361 GCCGGTCCCAAGATCCCATCACTCATGGGAGCGCCTGTAACAGGAGCGTACACTACT 420
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCAGAAACGTAGACGATTATGTCGTAAAATGATATGCGGCTCTTTTGGAGCG 480
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGTAAACGAGGCGGAGTACCGGCACAAATCAGTCTCTCTGTCACCGAAGATGCTATC 540
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTAGGGGCAACAGAAAACCTCGCGCCAGCTTTGGATCTTATGCTGACACATCAAT 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTAGCACAGTTTCTCTCCCGCGGACCTACAAAGGATGGAGCTATCAAAACCGACGTA 660
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 661 ATGCACACAGAACATTTATTTATCGGCAAGATCTTCTTTGGCTCCGGACTCCTCATTC 720
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 721 TGGSCAAACCATGACACAAATATGCTTATATGGTGGAAATCATCCATGGCGACACCGATT 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTAGCTGGTAACTGACAGTACGFGAACATTTTCATCAAAACAGAGAAATCACTCCT 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 841 AAACCATCTTCTGTAAGCAGCTCTTATTTGCCGAGCACTCATATCGCTCTTGGCTAT 900
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAGTGGAAACCAAGAGATGGGAAGAGTAACATGGACAAGTCACTTAATGTAGCTTTC 960
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 961 GTAAATGAAACAGCTCTTATCTACTACCAAAAGCTACGTATTCATTTACTGGCACA 1020
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1021 TCAGGCAAACTTTGAAGATTTTCATTCGTTGGTCTGATGCACCGGCAAGTACTCCGCA 1080
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCGGTTACATGGTGAATGATCTGGATCTGGTGATTACAGTCCAAATGGAAAGATAT 1140
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1141 GTTGGAAACGACTTTACTGCTCCCTATGATAATAACTGGGATGGACGTAACATGTAGAG 1200
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AACGTGTTTATCATGTCGCGAAAGCGGAACGTATACATGAGGTACAGGCTTACAAT 1260
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTTCCCAAGGCGCGAGGCGTTTCTTTGGCTATTGTGAAC 1302

RESULT 9
AR069954
LOCUS
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
```

```
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3003)
TITLE Sloma,A. and Christianson,L.
JOURNAL Nucleic acids encoding a polypeptide having protease activity
FEATURES Patent: US 5891701-A 41 06-APR-1999;
Location/Qualifiers
source 1..3003
/organism="unknown"
/mol_type="unassigned DNA"

Alignment Scores:
Pred. No.: 1,74e-125 Length: 3003
Score: 2120.50 Matches: 405
Percent Similarity: 97.91% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-1-COPY (1-434) x AR069954 (1-3003)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCGCGGCAATTTGTAAGACAGACGTCGCAAAATACTTTGGCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATATCGGCTTGATACAGGAAGAATGACAGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGTCATGAAGCATTCGCGGTAAAGATTACCCGACTATATGACCTGGGCAAGCAAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT-- 1706
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTCAAAATAAAGGATGGCACCCGACCAAGCCCAATCTAGTCTTTCAATCTATTATGGAT 1766
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGAGGCGCTGGGAGGACTACCTGCTTAATCTACAAATATTTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 1827 GCTGGAGCGAGAAATTCATACGAATTCATGGGCGCTCCAGTAAACGGTGCCTATACGACA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTTCGGGCC 1946
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAAATGAGGCGCCAGGTAGCGGTACAAATCAGTCACCGAGCAAGCAAAATAATGCGATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTGGGCGCAACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATTAATATTAAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCACGAGGCTCTACTAGAGATGGACGTATTAGCCGAGCGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGGTAGCTATATTTCTCTGCTAGATCATCATTAGCTCCAGATTCTCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
```

```

Db 2187 TGGGCAACCATCATAGTAATAATGCTACATGCTGGTGGTACTTCTATGGCTACTCCAATT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAATGTTGCACAATTAAGGAGCATTGTTGCAAAATAGAGGGTAACCTCCT 2306
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCTTTTAAAGCTGCTTTAATTCAGGTGCTGCGAGTTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGGTAACCAAGATGGGAGAGTAAGCTAGATAAATCCCTAAATGTGCAATT 2426
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTACGGCTCAA 2486
Qy 341 AlaGlyLysProLeuLysIleSerLeuValThrSerAspAlaProAlaSerThrThrAla 360
Db 2487 GCTGGTAACCCCTTAAATAATATCACTTGTGTCAGATGCCACCAAGTAGCACGGCA 2546
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCACCTAATCTTAGTGAATGATTAGACTTAGTAACTGTCACCAATAGGAACTAAATAC 2606
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 2607 GTCGGAATGACTTTACAGACCGTATGATAACAATGGCATGGCAGAGAACCAACGTGAA 2666
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGCTTTATCATGCTCTCTAAGCGGAGATATACAGTCGAAGTCGAGCTTACAAT 2726
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCGCAACCTTTCTTAGCGAATGTACAT 2768

RESULT 10
LOCUS BD062155 3003 bp DNA linear FAT 27-AUG-2002
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sloma,A. and Christianson,L.
Nucleic acids encoding a polypeptide having protease activity
Patent: JP 2001514529-A 39 11-SEP-2001;
NOVO NORDISK BIOTECH INC
PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PR 12-JUN-1997 US 08/873479
PI ALAN SLOMA,LYNNE CHRISTIANSON
PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..3003
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1.74e-125 Length: 3003
Score: 2120.50 Matches: 405

```

```

Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-1-COPY (1-434) x BD062155 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCCGTGGCATTGTGAAGCAGACGTCGCACAAATAACTTTGGCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGAACAAGGACAGATTGTAGCAGTTGCTACTGGCTTGGCTTGGATACAGAGAATAATGACAGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCATGTGATGAGCANTTCGCCGTAGATTACCGCACTATATGCACTGGCGCAGACCAAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCAAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT--- 1706
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATTAAGGATGGCACCGCAAGCAATCTAGTCTTTCAATCTATTATGGATAGT 1766
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCGTGGAGGACTTACCTGCTAATACAAACATTTATTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 1827 GCTGGAGCGAGATTATCATCAAAATCATGGGGGCTCCAGTAAACGGTGCCTATAGACA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTTCGGAATGTTGATGATTATGTGAGAAAAATGATATGACGATCTTTTTGGGCGC 1946
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
Db 1947 GGAATAGAGGACAGGACGCGTACAATCAGTGCACAGGACAGCAAAAATGCGATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCCGAAACCTACCTCCAGCTTCGGATCTTATGCGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCACGAGGCTCTACTAGAGATGGACGTATTAAAGCGGACGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCGAGTACGTATATTTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 2187 TGGGCAAAACCATGATAGTAATAATATGCTACATGGGTGGTACTTCTATGGCTACTCCAAT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAATGTTGCACAATTAAGGAGCATTGTTGCAAAATAGAGGGTAACCTCCT 2306
Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCTTTTAAAGCTGCTTTAATTCAGGTGCTGCGAGTTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGGTAACCAAGATGGGAGAGTAAGCTAGATAAATCCCTAAATGTGCAATT 2426
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340

```

Db 2427 GTGAATGAACGAGCCCTTTTCAACAGTCAAAAGCAACATATTCTGTTACGGCTCAA 2486

Qy 341 AlaclyysProLeuylsleSerLeuValTrpSerAspAlaProAlaSerThrAla 360

Db 2487 GCTGGTAAACCTTAAATAATACATCTGTTGTGATGATCCACGATGACGACGGCA 2546

Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValThrAlaProAsnGlyThrGlnTyr 380

Db 2547 TCACCTAACTTAGTGAATGATTAGACTTAGTAACTCACTGCACCAATGGAACATAATAC 2606

Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400

Db 2607 GTCGAAATGACCTTACAGCACCGTATGATGATGATGATGATGATGATGATGATGAT 2666

Qy 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrThrIleGluValGlnAlaTyrAsn 420

Db 2667 AATGTGTTTATCAATGCTCCTCAAGCGGAAGCTATACAGTCAAGTGCAGGCTTACAAT 2726

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Db 2727 GTACCAAGTACGCGCAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 11

AB046405

LOCUS

AB046405

DEFINITION

Bacillus sp. SD521 PROD gene for protease, partial cds.

ACCESSION

AB046405

VERSION

AB046405.1

GI:12381942

KEYWORDS

SOURCE

Bacillus sp. SD521

ORGANISM

Bacillus sp. SD521

REFERENCE

1 (sites)

Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.

AUTHORS

Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships

TITLE

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

JOURNAL

20568575

MEDLINE

11118284

REFERENCE

2 (bases 1 to 1299)

Saeki, K.

AUTHORS

Direct Submission

TITLE

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)

JOURNAL

Location/Qualifiers

FEATURES

source

1..1299

/organism="Bacillus sp. SD521"

/mol_type="genomic DNA"

/strain="SD521"

/db_xref="taxon:133780"

1..1299

/gene="PROD"

<1..>1299

/gene="PROD"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21268.1"

/db_xref="GI:12381943"

/translation="NDVARGIVKADVAQNNGYLGQGVVAVADTGLDTRDSSMHEAFRGKITALVALGRNTNANDPNGHTHVGSLGNALNGMAFQANLVFQSIMDSGG LGGLPSNLTLFSGAWNAGAR IHTNSWGA PVNGAYTANSRQVDEYVRNNDMTVLPAAG NEGNCSNTISAPGTAKNAITVGATENVRPSGLADNPNHIAOFSSRGATROGRIPD VTAPGTILSARSLAPDSSFANYNSKAYMGTSMTAPIVACNVQAUREHFIKNRG ITPKPSLIKALAGATDVGLGTFPSGQGMGRVLDKSLNVAIVNATATLQKATY SFQOAGKPLKISLVWDAPGSTASTYLVNDLDLVTAPNGQKYVNDPFSYPYDNNW DGRNNVENVFINAPQSGTYTIEVQAVNPSPQRFSLAIHV"

ORIGIN

Alignment Scores:

Pred. No.: 7,1e-118 Length: 1299

Score: 1993.50 Matches: 380

Percent Similarity: 94.24% Conservative: 29

Best Local Similarity: 87.56% Mismatches: 24

Query Match: 88.72% Indels: 1

DB: 1 Gaps: 1

US-09-985-689A-1-COPY (1-434) x AB046405 (1-1299)

Qy 1 AsnAspValAlaAArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1 AATGATGTAGCAGAGGAGTATGTAAGCAGACGTTTGCAAAACAATACGGACTATAT 60

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 61 GGACAAAGGTCAAGTAGTTGCGATAGCGGTTTAGATACAGGTCGTGAACGATAGT 120

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 121 TCTATGCATGAAGCATTCCTGGGAAATTAACAGCTCTTTACGCGTTAGGAAGAACTAAC 180

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 181 AATGCAATGATCCGAATGGGCATGGTACGCATGTAGCAGGTTCTGTACTTTGGTAAT --- 237

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 238 GCTTTTAAATAAAGGAATGGCTCCGCAAGCTAACTAGTCTTTCCAACTATTATTATGGATAGC 297

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

Db 298 AGCGAGAGTATTAGTGGATTACCATCGAATTTGAATCGTTATTATTAGTCAAGCTTGGAAAT 357

Qy 121 AlaGlyAlaArgGlyIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140

Db 358 GCTGGGGGTAGAAATTCATACTTCTGGGGTGTCTCAGTAAATGGAGCGGTACACTGCT 417

Qy 141 AspSerArgAsnValAlaAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 418 AACTCGACAGCAAGTGGATGATGTTTCCAAATTAATGATATGACGGTACTTTTTCGACGA 477

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

Db 478 GGTAAATGAAGTCCTAAATTCAGGAACAATATTAGTGCTCCAGGCACAGCAAAAATGCCATT 537

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 538 ACGGTCCGCGCAACAGAAACTATCGCCGAGCTTCGGTTTCATTAGCAGATAACCCAAAT 597

Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGACCGAAATTTAAACGACGTA 657

Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240

Db 658 ACAGCTCCTCGGAACAATTTATTATTCACGACGCTTCTTCCTTAGCCCCAGACTCTTCGTTT 717

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 718 TGGGCGAATTATACAGTAAGTATGCGTACATGGCGGTACTCTATGGCCGACACCTATA 777

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280

Db 778 GTTGGCGGGAATGTCCGCAATTTAGCGAGCATTTTATAAAAAATAGAGGAATACACCT 837

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

Db 838 AAACTTCTCTTAATAAGCTGCACCTTATCGCTGGGCTACTGATGTGGTTTAGGATAT 897

Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 898 CCAAGTCGTGACCAAGGCTGGGGCGCTGTACTCTAGATAAATCCTTAAATGTAGCGTAT 957

```

QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 958 GTCAATGAAGCAACGATAGCAACAGGACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1017
QY 341 AlaGlyLysProLeuLysLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1018 GCGGTAAACCTTTAAATATCTGTTAGTGGACAGATGACCTGGAGTACAACTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrGlnTyr 380
Db 1078 TCTTATACACTAGTTAAATGATTTAGATCTAGTTATTTAGTCTCCGAATGGCAAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyValGlnAsnValGlu 400
Db 1138 GTAGGAATGATTTTAGTTATCTTATGATATATACTGGATGGTGGCAAAATGTTGAG 1197
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1198 AACGTATTTATAACGCTCGCAATCTGGAACGTATACAAATGAGTTCAAGCGTATAAT 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1258 GTACCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 12
AB046402
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1. .1299
/gene="PROA"
<1. .>1299
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translational="NDVARGIVKADVAQNNVGLYQGQGVAVADTGLTGRNDSMHE
APRKITALYALGRNNDPNPHGTHVAGSVLGNALNKGMAPQANLVFQIMDSGQ
LGLPLSNLTLSQAWNAGARIHTNSWGANVYATANSROVDEYVRNNDMTVLFAAG
NEGPNSTGTSAPGTAKNALITVGATNENRPSFGSIADPNHIAQFSRRGATRDRIKPD
VTAPGFTILASRSLAPDSSFWANVNSKYVMGWTSMATPIVAGNVALRHHFTKNRG
ITPKPSLIKALIAAGTAVGLYPSGDGQWGRVTLDSLNVAIVNEA TALTTGOKATY
SFQTQAGKPLKISLWNTDAPSGSTTASYTLVNDLDLVIITAPNGQKIVGNDFFSYPDNNW

```

```

ORIGIN
Alignment Scores:
Pred. No.: 1,27e-117 Length: 1299
Score: 1989.50 Matches: 379
Percent Similarity: 94.0% Conservative: 29
Best Local Similarity: 87.33% Mismatches: 25
Query Match: 88.54% Indels: 1
DB: 1 Gaps: 1
US-09-985-689A-1-COPY (1-434) x AB046402 (1-1299)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTAGCAAGCAAGATAGTAAAGCGACGTTGCAAAACAAATACGACTATAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGCAACAGGTCAGTAGTTCAGTAGCGGATACGGGTTAGATACAGGTCGTAAACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGCATGAGCATTCCTCGTGGGAAATTCAGCTCTTTACGCTTAGAGAGAACTAAC 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCAAAATGATCCGAATGGCATGTGTCGCTAGCTGTTCTGTGCTTGGTAAT--- 237
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCTTTAAATAAGGAATGGCTCGCAAGCTAACTTAGTCTTCCAAATCTATTATGGATAGC 297
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 298 AGCGGAGGATAGGAGGATACCATCGAATTTAAATACGTTATTAGTCAAGCTTGGAT 357
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 358 GCTGGCGCTAGAAATTCATCACTAATCTCGGGGGCCCCAGTAAATAGGACGCTACTGCT 417
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 418 AACTCGAGACAGTGGATGAGTATGTCGAACATGATATGACGGTACTTTTTCAGCT 477
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGAATGAAGACCTTAATCTCGGAACATATAGCGCTCCAGGGACAGCGAAAATGCCATT 537
QY 181 ThrValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACAGTGGCGCAACGGAACAACTACCGACCAAGTTTGGTTCAATTGCAGATTAACCTAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 598 CATATCGCACAAATTTTCATCGAGGAGGACTACGAGAGATGGACGAATTAACCCAGACGTA 657
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGTCTCTGGAACATTTATTTATCAGCACGCTCTCTTTAGCACGACACTCTTCGTTT 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 718 TGGGCGCAATTTATAACAGTAAGTATGCTATATGGGCGGTACCTCTATGCGCACACCTATA 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTGGCGGGGAATGTGCGCAATTTACCGGAGCATTTTATAAAAATAGAGAAATACACCT 837
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 838 AAACCTTCCTTTAATAAAGCTGCACATTATCGCTGGGCGCTACTGATGTTGTTTAGGATAT 897

```



```

Qy 301 ProAsnGlyAsnGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaThr 320
Db 898 CCAAGTGGTGAACAGAGCTGGGGCGGTGTTACTTTAGATAAATCGTTAAATGAGCGTAT 957
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 958 GTCAATGAAGCAACTGCATTAACAACAGGACAAAGCAACGATTTCGTTCCAAACGCCAA 1017
Qy 341 AlaGlyLysProLeuLysLysSerLeuValTTPSerAspAlaProAlaSerThrAla 360
Db 1018 GCGGGTAAACCATTAATAAATCTCGTTAGTAGGACAGATCCACCTGGAAGTACACAGCA 1077
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValThrAlaProAsnGlyThrGlnTyr 380
Db 1078 TCTTATACACTAGTTAATGATTAGATCTAGTTACTTACTCTCGAATGACAAAATAT 1137
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTTPAspGlyArgAsnValGlu 400
Db 1138 GTAGGTAATGATTAGTTATGCTTATGATAAATATGGATGCTGCAACAAATGTTGAG 1197
Qy 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaThrAsn 420
Db 1198 AACGTATTTATAACGCTCCGCAATCTGGAACGTATACATTCAGGTTCAAGCGTATAAC 1257
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1258 GTTCCATCTGACCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 13
AB046404 1299 bp DNA linear BCT 23-JAN-2001
LOCUS
DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
ACCESSION AB046404
VERSION AB046404.1 GI:12381940
KEYWORDS
SOURCE Bacillus sp. Y
ORGANISM Bacillus sp. Y
REFERENCE 1 (sites) 1299 bp DNA linear BCT 23-JAN-2001
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki,K.
TITLE Direct Submission
SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan E-mail:387185@kstanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
Location/Qualifiers
1..1299
/organism="Bacillus sp. Y"
/mol_type="genomic DNA"
/strain="Y"
/db_xref="taxon:133779"
1..1299
/gene="PROC"
<1..>1299
/gene="PROC"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="protease"
/protein_id="BAB21267.1"
/db_xref="GI:12381941"
/translation="NDVARGIVKADVAQNNGYLYGQGVVAVADTGLDTRNDSSMHE
AFRKIKTALVALGRTNNSDPNGHGVHAGSVLGNALNKGMAPQANLVFQSDMSGG
LGGLPSNLNTLFSQANNAGARIHTNSKGPVNGAYTANSRQVDEYVRNDMTVLPFAG

```

ORIGIN

Alignment Scores:

Pred. No.: 3,55e-117 Length: 1299
 Score: 1982.50 Matches: 378
 Percent Similarity: 94.01% Conservative: 30
 Best Local Similarity: 87.10% Mismatches: 25
 Query Match: 89.23% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-1-COPY (1-434) x AB046404 (1-1299)

```

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTAGCAGAGGATAGTAAAGCTGATGTTGCACAAAACAATACGGATTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAAGTCAAGTAGTTCAGTAGCGGACACAGGCTTAGATACAGGTCGTACGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGATGAAGCATTCGCGGGAATACACGCTCTTTACGGCTTAGGAAGAACAATAA 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGGAGTGTATCCGATGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAAT 237
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAAATCTATTATGGATAGC 297
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 298 AGCGAGGATTAGTGGTGTACCATCGAATTAATAGTATTATTAGTCAAGCTTGAAT 357
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTPGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 358 GCTGGAGCAGAAATTCATACATACTCTTGGGAGCCCGAGTAAATGGAGCGTACACTGCT 417
Qy 141 AspSerArgAsnValAlaAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 418 AACTCGACACAAGTGGATGAATATGTCGAAATATGATATGACGGTACTTTTTCAGCT 477
Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGTAATGAAGTCTCTTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAAAAATGCTATT 537
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACGGTCCGCGCAACGGAATACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAAT 597
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 598 CATATTGCACAATTTTCATCCAGAGGAGTACAGAGGATGGACCAATTAAGCCTGACGTA 657
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGCTCCTCGGAACAATTTATTTATCAGCACGTTCTTCTTCTTAGCTCCAGACTCTTCGTT 717
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 718 TCGCGCAATTTATACAGTAATATCGGTATATGGCGGTACTCTCATGGCGACACCTATT 777
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTCAGGGAATGTCCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCT 837

```

NEGPSGTISAPGTAKVAITVGATENYRPSFGSIADNPNIHQFSSRGATRDGRIKPD
 VTASGTILSARSLAPDSSPMYNSKYVMGTSMTATPIVAGNVLAQLRHFHKRG
 ITPPSILKALAGATDGLGYPNGQGWRTYLNKSLAVYNEATLATGSKATY
 SFQAGKPKLSILWTFADPGSTTASTILVNDLVLVTAPNGQKYVGNDFSPYDNNW
 DGRNNVENFINPQSGTYIIEVOAYNVPSQPQFSLAIHV*


```

Db      1721 CCTAGTGGTACCACGAGCTGGGCGGTACTCTAGATAAATCGTTAAATGAGCGTAT 1780
Oy      321 ValAsnGluSerSerSerLeuSerThrSerGlnIysAlaThrTySerPheThrAlaThr 340
Db      1781 GTCAATGAAGCAACTGCATTAGGCACAGGACAAAGCAACGATTCGTTCCAAAGCACAA 1840
Oy      341 AlaGlyIysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db      1841 GCGGTAACCTTTAAAAATCTCGTTAGTAGGACAGATCTCTCTGGAAGTACAACGTGCA 1900
Oy      361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db      1901 TCTTATACACTAGTAATGATTAGATCTAGTCTAGTTACTCTCCGAATGGACAAATAAT 1960
Oy      381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGlu 400
Db      1961 GTAGGAATGATTAGTATCTCTATGATAATAAATCGGATGGTGGCAACAATGTTGAG 2020
Oy      401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db      2021 AACGTATTATTAAACGCTCGCAATCTGGAACGTATATATATGAGGTTCGAAGCGTATAAT 2080
Oy      421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db      2081 GTACCAATCTGCCCAACGCGTTTCTCACTAGTATCGTACAT 2122

RESULT 15
AF268611/c 60006 bp DNA linear BCT 30-MAY-2001
LOCUS      37f11, partial sequence.
DEFINITION uncultured marine group II euryarchaeote EBAC37f11, BAC clone
ACCESSION AF268611
VERSION    AF268611.1 GI:9664575
KEYWORDS   uncultured marine group II euryarchaeote 37f11
SOURCE     Archaea; Euryarchaeota; Marine Group II; environmental samples.
ORGANISM   Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
            Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
            Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
            Construction and analysis of bacterial artificial chromosome
            libraries from a marine microbial assemblage
            Environ. Microbiol. 2 (5), 516-529 (2000)
JOURNAL    21128453
MEDLINE    11233160
REFERENCE  2. (bases 1 to 60006)
AUTHORS    Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
            Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
            Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
            Direct Submission
            Submitted (17-MAY-2000) R. & D. Monterey Bay Aquarium Research
            Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES   source
            1..60006
            /organism="uncultured marine group II euryarchaeote 37f11"
            /mol_type="genomic DNA"
            /db_xref="taxon:133822"
            /clone="BAC 37f11"
            /note="environmental sample"
            1484..3178
            /note="37f11#1"
            /codon_start=1
            /evidence=not experimental
            /transl_table=11
            /product="S-layer protein"
            /protein_id="AAF97179.1"
            /db_xref="GI:9664576"
            /translation="MEPQTELDIFVQSPSSAATAGQNVIKRVDSGCGGLEEEFI
            PVRVSPNFELASQNFVSEGGYGLAWMENTGMTSHVTITLENLPGVNIISGMQ
            SASLSPHARIPPEIIPPTNWDKQPLEIISIVHPQMGVSEITVIHNSIFASTP
            VLSGSGSSQAFSLHLSINPSTVTVPDATNQDEYRLILEEGINPETISTNDIAIQ

```

```

UNIGSILPQVSACQFISDVQFQELGRAPLTETVWSCSVNGDLDEQTLTFTIASTNTGE
TIPDLGFTIQRNESFVNISVDQWPEPQQLIRVVYDNGGIEVVSIEIAQVSRO
SNWVGIASFATGDLDIATRTNYEVLGVNVCILITVTSSTSVSRIERIIVDVEGQPA
PIVKIOPNSIDRESLSATIDCNPSFVDDDDPNTATALYVESQSILVTTNNLLMET
AITTIFGAYLVYVQRRRAGVQENIRSKPKPTNNKTQETKETEYVEDDMSLIEE
EMEUEEPVSLVEEITVSETDLSPSGRLDITIRQELDPDVEIVDTISIEERMSKFFD"
3223..3891
/note="37f11#2"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/db_xref="GI:9664577"
/translation="MQNSNSTSAFYTLDCGVSENLDETFOALLEIRGNHQLNEKLV
NREAFSTLKKQGLDEIDDRNKITWDDATLFTCRMLNRDSAPIMLGNKNRFRGFR
PHFTGSLQVILENLSIPQSEGLYSDLIEHLEYLTERCSEMTGKHEFLEGKAGL
MLGVITFEYKALRSLMLGSGVMVSRREPLDGGVREAVRHLNALLMAAERRGAGLIH
RQHE"
complement(3892..4797)
/note="37f11#3"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="SAICar synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translation="MGENLYQGKVKQVWSTDDPDLLEFRFTNQISVFDQIIPSLIPRK
GESNRTTAHFKLVEQGVGCTHLVELNAPDRCLVRKVEVFKPGPAIDRAEWVFP
LEFIVRVLGSAWRFORFORGLVSLGSDAEYGVKLSFPFVETVTKFEAFDRNID
REELQISNITDQFDSIQVAAKLVDEIIEEAAKNGLIHVDGKKEFALGPKGVLV
DTFGTLDLRWDAAEYNGECIELSKFVTRTHVINSGHQAEKLARDGSAIDPPIPA
LPESVIEETASLYSMYERLTSQQF"
complement(4846..5508)
/note="37f11#4"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translation="MPDITLDIHGSDQIGIHLSVGNVLFHPRELPTTIEKLQVVG
LELAPISIGGNLIGALAGNTKGMADIATESDIDALTAIGDVVWEGVNTAGND
LVNVEGCVASPSIPKDLGLDILSEVLGDIITSTTIGQDVVGSVGNVSGVLLHPDV
APEVLLIEILGVPVPMVGTVSGPYVGAGICASNGALAGSETTGPENLRIDALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..5811
/product="23S ribosomal RNA"
complement(8888..10192)
/note="37f11#7; contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translation="MNECAVCGILYLAGSPCAGCQLRAQDQSPSDAAALPTEVPG
LDAAQAWYDLEGTEAPVEQEEAASLPFGGEGSOTNIRLPGIGSHPEGIPFEL
SNODSGQEEVVOHVEPIESTNLFPVALETCHDPSVYASVEIEPVLAPVEAD
LPLEIEPVQVPVPGPIRLTALPFPVEPTVDSLPEWMSIESFGSGDQIFTAEEDVIE
VVFSDLETVVHRHGDENYPMERKDSVYSKSLQPFDLHPAQLMVDVGTDSARAL
LESQFTSIGQDWRSAARSPQIVARFPQDSGAMNNGYIALQLQVAVNQHSNDIEVS
NSATQFEAAIILREAVTNSQSPESYINLSQALLSGREKALLDLMAGNESQRES
NFTNLRAAILAQLGRYDEAKLLVLPQADKLASGNLLKLPAL"
complement(10248..11969)
/note="37f11#8"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"

```



```
Qy 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 277
Db 16658 ACAGCAGTACGAGTGGTTCAGCATCCCTCGCAGCTGAATATCTGCGTGAAGTTGCTGGA 16599
Qy 278 IleThr---ProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIle 296
Db 16598 ATCAACAACCATCTGCAAGCCCTATTAGGCAACTCTCATCAACGGGGCAGAGATTTA 16539
Qy 297 GlyLeu---GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSer 315
Db 16538 GGCACACGGACATTCCTAACCAACGAGGATGGGGCAAAATGATCTCGAAATAGC 16479
Qy 316 LeuAsnValAla-----TyrValAsnGluSerSerLeu 327
Db 16478 CTGAAATCTTCGTCGTCGTCGGCTTCCTCGAGTCTTTTCAGGATGATGAGCGTGAACIT 16419
Qy 328 SerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIle 347
Db 16418 CAGCAGGATTCCTCATTCGATTATTCTCTTTCACCTTGATGATCAAAGGGTATTGACATT 16359
Qy 348 SerLeuValTTPSerAspAlaProAlaSerThrThrAlaSerValThr-----Leu 364
Db 16358 ACCTTGGCATGGAGCGATGCAAGCCAGTCGCAATCTGCAATCTGAATCAGCTCTA 16299
Qy 365 ValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAsp 384
Db 16298 TTGAACAATTTAGATTGATTCTTAATTGCACCCGATGGGTCTTCATATTGGGTAAATGAC 16239
Qy 385 PheThrSerProTyrAsn-----AspAsnTrpAspGlyArgAsnAsnValGluAsn 401
Db 16238 TTCCTTCTGGGATTTCACAACACAGGGGTTCGGCTGATTAATCTGACAATATCGAGCGC 16179
Qy 402 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluVal----- 416
Db 16178 ATTGCAATCCCTGCGAGGCGACGACACAGCGTGTGATTGGATGGTTACAGTTGAACAT 16119
Qy 417 -----GlnAlaTyrAsnValProVal 423
Db 16118 CGAGGAGGAGTTCTCAGCGCTACAGTATCGTTATT 16083
```

Search completed: March 15, 2004, 21:48:31
Job time : 3658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 19:16:41 ; Search time 395 Seconds
(without alignments)
4667.640 Million cell updates/sec

Title: US-09-985-689A-1-COPY

Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPOTFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=vlh
-Q=/cgn2.1/USPTO.spool/SWOPR985689/runat_10032004.112806.19350/app_query.fasta_1.583
-DB=N_Geneseq_29Jan04 -QPM=faetap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=SWOPR985689 @CEN 1.1 352 @runat_10032004.112806.19350 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2242	99.8	1923	2 AAX37278	Aax37278 Bacillus
2	2237	99.6	1923	2 AAX37279	Aax37279 Bacillus
3	2178	96.9	1920	2 AAX37277	Aax37277 Bacillus
4	2120.5	94.4	3003	2 AAV82382	Aav82382 Bacillus
5	1981.5	86.2	1293	2 AAQ27515	Aaq27515 Alkali-pr
6	447.5	19.5	1977	2 AAT85667	Aat85667 Thermococ
7	447.5	19.9	1977	2 AAX05926	Aax05926 WO9856926
8	411.5	18.3	1236	2 AAX05920	Aax05920 Hyperther

9	411.5	18.3	1566	2 AAT85668	Aat85668 Pyrococcus
10	411.5	18.3	1962	2 AAT85695	Aat85695 Pyrococcus
11	411.5	18.3	1962	2 AAX05929	Aax05929 Hyperther
12	398.5	17.7	1977	2 AAT85669	Aat85669 Protease
13	372	16.6	2121	4 ABL54900	Ab154900 T. yonsei
14	348	15.5	2539	2 AAT61454	Aat61454 Streptomy
15	348	15.5	2809	2 AAT61455	Aat61455 DhpA-mel
16	336	15.0	135638	7 ABX34289	ABX34289 S. atrool
17	332.5	14.8	1329	6 ABK74643	ABK74643 Bacillus
18	311.5	13.9	2532	2 AAQ29134	AAQ29134 Encodes R
19	307	13.7	2835	2 AAT08141	Aat08141 Hyperther
20	307	13.7	4765	2 AAT08132	Aat08132 Protease
21	307	13.7	4765	2 AAT85670	Aat85670 Pyrococcus
22	307	13.7	4765	2 AAX05921	Aax05921 WO9856926
23	306.5	13.6	1560	7 ABZ37569	ABZ37569 Streptomy
24	306.5	13.6	59816	7 ABZ37516	ABZ37516 Streptomy
25	306.5	13.6	59816	7 ABZ37515	ABZ37515 Streptomy
26	302.5	13.5	1859	2 AAT85677	Aat85677 Thermococ
27	296.5	13.2	3413	2 AAV72330	Aav72330 F. balust
28	296	13.2	898	2 AAT08131	Aat08131 Hyperther
29	290.5	12.9	564	2 AAT08134	Aat08134 DNA seque
30	290.5	12.9	564	2 AAT85676	Aat85676 Thermococ
31	282	12.6	3788	9 ADD24905	Add24905 DNA encod
32	278	12.4	3743	9 ADD24901	Add24901 DNA encod
33	276	12.3	1306	6 ABL55784	AB155784 Bacillus
34	276	12.3	1330	6 ABL55787	AB155787 Bacillus
35	275	12.2	546	4 ABL53453	AB153453 T. yonsei
36	270.5	12.0	2273	2 AAQ04339	Aaq04339 Aquaricin
37	270.5	12.0	2274	2 AAQ12838	Aaq12838 Aquaricin
38	270.5	12.0	2274	2 AAQ75859	Aaq75859 Aqualysin
39	261	11.6	4716	9 ABQ80437	ABQ80437 CspA codi
40	261	11.6	4740	7 ABZ58957	ABZ58957 Group B S
41	256.5	11.4	1152	8 ACC85473	Acc85473 B gibsoni
42	256.5	11.4	1194	8 ADA32120	Ada32120 DNA encod
43	256.5	11.4	10215	2 AAT39279	Aat39279 Transposo
44	256	11.4	1485	6 ABK74647	ABK74647 Bacillus
45	254	11.3	4650	6 ABN71526	ABn71526 Streptoco

ALIGNMENTS

RESULT 1

AAX37278

ID AAX37278 standard; DNA; 1923 BP.

XX

XX AAX37278;

AC

XX

DT 20-MAR-2003 (revised)

DT 21-JUL-1999 (first entry)

XX

DE Bacillus alkaline protease encoding DNA.

XX

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;

KW washing composition; oxidising agent; ss.

OS

OS Bacillus sp.

XX

PN WO9918218-A1.

XX

PD 15-APR-1999.

XX

PF 07-OCT-1998; 98WO-JP004528.

XX

PR 07-OCT-1997; 97JP-00274570.

XX

PA (KAOS) KAO CORP.

XX

PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

PI Shikata S, Nomura M;

XX

DR WPI: 1999-287736/27.

DR P-PSDB; AAY17088, AAY17090.

XX

PI Shikata S, Nomura M;
 XX WPI: 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 PS Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,24e-162 Length: 1923
 Score: 2237.00 Matches: 432
 Percent Similarity: 99.77% Conservative: 1
 Best Local Similarity: 99.54% Mismatches: 1
 Query Match: 99.55% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AAX37279 (1-1923)

QY 1 AsnAspValAlaAArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 619 AATGATGTGGCGGTGGAATGTCAAGCGGATGTGGCTCAGACGACGTACGGGTTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAAGACAGATCGTAGCGGTGGCCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGCATGAAGCCTCCCGCGGGAATAATCTCATTTATATGATTTGGACGGACGAAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGGTCATGGTACGCATGTGGCTGGCTCCGTATTAGGAAACGGC 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTAATAAGGAATGGCCCTCAGCGCAATCTAGTCTCCAAATCTATCATGGATAGC 918
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 919 GGTGGGGACTTGAGGAGTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 979 GCTGGTGCAGAAATTCATACAAATCTCCGGGAGCAGCAATGGGCTTCACAAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCITTCGTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 CGGAATGAAGACCGAAGCGGGAACCATCATGTGCACGGCAGCAGCTAAATAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

DB 1159 ACAGTCGGAGCTACGGAAAACTCCGCCCAAGCTTGGTCTTATCGGACAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleLysProAspVal 220
 DB 1219 CATGTGGCACAGTCTCTTCACGTGGACCGAAGGATGACGGATCAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1279 ATGGCACCGGAAACGTTTATATCTATCAGCAAGATCTTCTCTTGACCGGATTCCTCTTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGGGCAACCATGACAGTAATAATATGCAATATGTTGGTGGAAAGTCCATGGCTACCCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1399 GTTGTGGAACGTCGGACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1459 AAGCCTTCTTATTAAAAAGCGCAGCTGATTCGGGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1519 CCGAAGCTTACCAAGGATGGGACGAGTGACATTGGATTAATCCTCGACGTTGCTAT 1578
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAGGACGCTACTCTGTTTACTGCTACT 1638
 QY 341 AlaGlyLysProLeuLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 DB 1639 GCCGCAAGCCTTTGAAATCTCCTCGTATGGTCTGATGCCCTCGCAGCACAACCTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTTCATTACCGCTCCAAATGGCACACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAATGACTTTACTTCGCCATPACATGATTAATCTGGGATGGCCGCAATAACGTAGA 1818
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 1819 AATGATTTATTATTAATGACCAACAAAGCGGACGTATACATTAAGTACAGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGTTGGACCAAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX
 AC AAX37277;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR P-PSDB; AAY17087, AAY17089.
 XX
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4, 25e-158 Length: 1920
 Score: 2178.00 Matches: 417
 Percent Similarity: 99.31% Conservatives: 14
 Best Local Similarity: 96.08% Mismatches: 3
 Query Match: 96.93% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689a-1-COPY (1-434) x AAX37277 (1-1920)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 616 AATGATGTGCCAGAGTATTGTCAAGCGGATGTGGCAGACAGACGACTACGGTTTGAT 675
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GCACAGCGCCAGATTGTCGCGATGCGGATGCGATGCGATGCGATGCGATGCGATGCG 735
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGCATGAAGCTTCGCGGTAAATAAACAGCACTATATGCACTGCGGTGCGACCAAT 795
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 796 AATCGAATGATACGAACGTCATGGTACCGATGCGGATGCGGATGCGGATGCGGATGCG 855
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 856 GCAACGAATAAAGGAATGGCACCTCAAGCAATCTGGTTTTTCAATCCATCATGATGAC 915
 Qy 101 GlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 916 AGTGGTGGGTGGAGGCTTGGCTTCCAACTCTGCAAACTTATTCAGCCAAAGCAATTCAT 975
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
 Db 976 GCAGTCCAGATTTCATCAAACTCTCTGGGGGCGAGCGGTGAATGGGCGCTACACGACA 1035
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCCAGAAATGTGGATGACTATGTAAAGGAAATGATATGACGATTTCTTTTCGGCGCT 1095

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1096 GGGATGAAGCGCGAAGCGGGTACCATCAGTGCACCTGTACGGCTAAAGACGCCATA 1155
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTCCGGCGCAACCGGAAACCTGCGTCCAAAGTTCGGTTCCCTATGCGATAATATTAAAC 1215
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1216 CACGTTGCACAGTTCTCTTCCCGTGGCCGACAAAAGATGGCGAATCAAGCCGTGATGTC 1275
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCGAGGACATATTTTATCAGCAAGATCTTCTCTGCACCCGATTCCTCCTTC 1335
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGGATCATGACAGCAATATGCTATATGGGTGAACGTCCTCATGGCAACACCGATT 1395
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1396 GTTCGGGGAATGTGCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGCAATCACTCCT 1455
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1456 AAGCTTCCCTATTGAAAGCAGCTTGAATTCAGAGTCTGCTGATGTGTGATGGGTAT 1515
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACGAAACCAAGGATGGGCGGAGTGACCCCTGGATAAATCGTTGAAACGTTGCCAT 1575
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1576 GTGAACGAATCCAGTGGCTTCACTCACTAGCCAAAGACACATATACCTTTACTGCAACG 1635
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1636 CGCGGCAAGCCATTGAAATCTCCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1695
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1696 TCTGTAAACCCGTGCAATGATTGGATTTGGTCAATTACAGCAACCAAGCAAGATAT 1755
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1756 GTGGGATGACTTCTCAGCACCATTTGACATACTGGGATGGCGCCCAATACGTAGAA 1815
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1816 AATGTATTATTAAATTCGCCCAAGTGGAACCATATACCATTTGAGGTGCAAGCATATAAT 1875
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1876 GTGGCGGTGGACCAACAAACTTCTCGTTGGCAATTCGTGAAC 1917

RESULT 4
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE *Bacillus* JP170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 XX flavour; ss.
 OS *Bacillus* sp.
 XX
 XX Key Location/Qualifiers
 FH 846..2771
 FT CDS

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 XX
 AC AAQ27516;
 XX
 DT 05-FEB-1993 (first entry)
 XX
 DE Alkali-protease Ya enzyme gene.
 XX
 KW Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 OS Bacillus sp. Y.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1299
 FT /*tag= a
 XX
 XX JP04197182-A.
 XX
 XX 16-JUL-1992.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX (LLOY) LION CORP.
 XX
 XX WPI; 1992-288440/35.
 XX
 XX P-PSDB; AAR26274.
 XX
 XX DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 XX resistance and improves detergency.
 XX
 XX Claim 3; Page 2; 17pp; Japanese.
 XX
 XX The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 XX
 XX Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 3,41e-143 Length: 1299
 Score: 1981.50 Matches: 378
 Percent Similarity: 94.01% Conservative: 30
 Best Local Similarity: 87.10% Mismatches: 25
 Query Match: 88.18% Indels: 1
 DB: 2 Gaps: 1
 US-09-985-689A-1-COPY (1-434) x AAQ27516 (1-1299)
 Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATAGCAGAGGGATAGTAAAGCTGATGTTGCACAAAACAACTACGGATTATAT 60
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyAAspAspSer 40
 Db 61 GGACAAGGTCACACTAGTTGCGAGTAGCGGACACAGGCTTAGATACAGGTCGTACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCATTCGCGGGAATAACACAGCTCTTTACCGGTTAGGAAGAACTAAT 180
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATCGAGTGTCCGATGGCATGGCACATGATAGCAGGTTCTGTACTTGTGTAAT --- 237
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAACTATTATGGATAGC 297

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACCGGAGGATTAGTGGCTTACCATCGAATTAATACGTTATTATTAGTCAAGCTTGAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 358 GCTGGAGCAAGATTACATACTTCTGGGAGCCCCAGTAAATGGAGCGTACACTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGACAAGTGGATGAGTATGTTCCGAATATGATATGACGGTACITTTTTCAGCT 477
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAATGAAGTCTCTAATTTCAGGAACAATAGTCTCCAGGTACAGCGAAAAATGCTATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACCGTGCGGCGCAACGAAACTATCGCCCAAGCTTCGGTTCCGATAGATTAACCCAAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGAGCAATTAAGCCTGACGTA 657
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerIleuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCCTGGACATTTAATTTATCAGCAGCTTCTTCCTTAGCTCCAGACTCTTCGTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGAATTATAACAGTAATACGCTATATGCGCGGTACCTCCATCGGCGACACTATT 777
 Qy 261 VallalaglyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTCAGGGAATTCGCGCAATTTACGTGAGCAITTTATAAAAAATAGAGTATTACTCCT 837
 Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAGCTTCTTTAATAAAGCTGCACTTATCGCTGCTACTGATGTTGGTTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCTAGTGTGCCAACGCTGGGCGCTGTTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 Qy 321 VallanGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAAGCACTGCTATGACGACAGGACAAAAGCAACGATATTGTTCCAGGCAAA 1017
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCTTTAAAAATCTCGTTAGTAGGACAGATGCTCTCGGAAGTACAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACATAGTTAATGATTAGTATTAGTATTACTGCTCCGATGGCAAAAATAT 1137
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1138 GTAGGAATGATTTTAGTTATCTTATGATAAATAACTGGGATGCTCGCAACAATGTTGAG 1197
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGATTATTAAACGCTCCGCAATCTGGAACGATATATAATTGAGTTCGAAGCGTATAAT 1257
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTACCATCTGCCCCACAGCGTTTCTCACTAGCTATCGTACAT 1299
 RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Thermococcus protease coding sequence.
 XX Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX Thermococcus celer; DSM-2476.
 XX WO9721823-A1.
 XX 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP003253.
 XX 12-DEC-1995; 95JP-00323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 XX Tsunawasa S, Kato I;
 DR WPI; 1997-332794/30.
 DR P-PSDB; AAW24121.
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 3; Page 86-87; 159pp; Japanese.
 CC This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred No.: 4.76e-25 Length: 1977
 Score: 447.50 Matches: 137
 Percent Similarity: 44.66% Conservative: 68
 Best Local Similarity: 29.85% Mismatches: 153
 Query Match: 19.92% Indels: 101
 DB: 2 Gaps: 18

US-09-985-689A-1-COPY (1-434) x AAT85667 (1-1977)

Qy 8 ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyGlnIleValAla 27
 Db 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGGCTACGACGAAAGCGGTGTGGTTGCC 492
 Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCTGATACGGGTATAGACGGGAC-----CACCCCGATCTGGAAG 534
 Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
 Db 535 GGCAAGGTCATAGCTGCTGATACGACCGCTCAACGGCAGGTCGACCCCTACGATGACCAG 594
 Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
 Db 595 GGACACGGAAACCCACGTTGCGGGTATCGTTGCCGGAACCGCGACGCTTAACCTCCAGTAC 654
 Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 Db 655 ATAGCGCTGCCCCCGCGCGAAGCTCGTGGGTCAGGTTCTCGGTGCGGACGGTTGCG 714
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121

Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTCCAGAACAGGACAAGTAC 774
 Qy 122 GlyAlaArgIle-----HisThrAsnSer 129
 Db 775 GGGATAAGGTCATCAACCTCTCTCCGCGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
 Qy 130 TrpGlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAspTyVal 149
 Db 835 CTCAGTCAGGCGCTCAACACGCTGGGACGCC----- 867
 Qy 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
 Db 868 -----GGTATAGTAGTCTCGGTCCGCGCAACGCGGCGCAACACCTACACC 918
 Qy 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTCCGCTCACCGCGCGCGAGCAAGGTATACCGTCGGTCA----- 963
 Qy 190 ProSerPheGlySerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgly 209
 Db 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGCAGGGA 1005
 Qy 210 ProThrLysAspGlyArgIleLeuPheProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1006 CCGACCGCGGACGGAAGCTCAAGCGGAGTCTGCGCCCGCGGTGACATCATAGCC 1065
 Qy 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAla 249
 Db 1066 CCGCGCGCCAGC-----CGAGCTGGACCCCGGACAGGTGAAG 1212
 Qy 257 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 Db 1213 ACCGCCCTCATCGAGACCGCGACATAGTCGCCCAAGGAGATAGCGACATCGCTAC 1272
 Qy 299 GlyTyProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
 Db 1273 GGTGCG-----GGTAGGTGAACGTCTACAAGGCCATCAAGTAC 1311
 Qy 317 ---AsnValAlaTyValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTy 335
 Db 1312 GACGACTACGCGCAAGCTCACCTTCACCGCTCCGCGCGACAGGGAAGCGCCACCAC 1371
 Qy 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 Db 1372 ACCTTCGACGTCAGCGCGCCACCTTCGTGACCCGCCACCTCTACTGGGAC----- 1422
 Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuValIleThrAlaPro 375
 Db 1423 -----ACGGGCTCGAGCAGCATCGACCTTACTCTAGACCCC 1461
 Qy 376 AsnGlyThrGlnTyValGlyAsnAspPheThrSerProTyAsnAspAsnTrpAspGly 395
 Db 1462 AACGGGAACGAG-----GTTGACTACTCTACACCGCTACTAC----- 1500
 Qy 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGlu 415
 Db 1501 -----GGCTTCGAAAGTCCGGTACTACACCCGACCGCGGAACCTGGACGGTCAAG 1554
 Qy 416 ValGlnAlaTyAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1555 GTCGTCAGCTACAAG-----GGCGCGCGCAACTACCAGGTCGACGTCGTGCTCAGC 1602

RESULT 7
 AAX05926
 ID AAX05926 standard; DNA; 1977 BP.

QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1555 GTGTCAGCTACAAG-----GGCGGGCGGAACCTACCAGGTGCAGCTGTCAGC 1602

RESULT 8
AA05920
ID AAX05920 standard; DNA; 1236 BP.
XX AAX05920;
XX 06-MAY-1999 (first entry)
XX Hyperthermostable protease fragment encoding DNA.
XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX Pyrococcus furiosus.
XX W09856926-A1.
XX 17-DEC-1998.
XX 04-JUN-1998; 98WO-JP002465.
XX 10-JUN-1997; 97JP-00151969.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
XX WPI; 1999-080907/07.
XX P-PSDB; AAW94836.
XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
PT gene encoding it, for large scale production of the protease for
PT industrial use.
XX Claim 6; Page 37-38; 82pp; Japanese.
XX The invention relates to a hyperthermostable protease derived from a
CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
CC 95 deg.C. The invention also provides gene sequences encoding a
CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
CC peptide from subtilisin, and PRO is the above protease. Host cells
CC (especially Bacillus strains) transformed with vectors comprising the
CC genes are used for the recombinant production of the protease. The
CC hyperthermostable protease which can be prepared in quantity suitable for
CC industrial use, can be used as an additive for drugs, washing agents and
CC foodstuffs and for chemical synthesis
XX SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.59e-22 Length: 1236
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 2 Gaps: 20

US-09-985-689A-1-COPY (1-434) X AAX05920 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
DB 24 GTCTGCAGCTCAAGTATATGGCACTTACGTTGGAACTTGGGATATGATGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
DB 84 CACAATAGGAATAATTGACACTGGAAATTGAC-----GCTTCTCATCC 123

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
DB 126 AGATCTCCAGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 170

QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
DB 171 TAGGAGTTATCCATACGATGACCATGAGCATGGAATCCTGAGTCTCAATAGCAGCTGG 230

QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
DB 231 TACTGGAGCAGCAAGTAATGCAAGTACAAGGGAATGGCTCCAGGAGTAACTGGCGGG 290

QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
DB 291 AATTAAGGTTCTAGGTCGCGATGCTTCTGGAGCATATCTACTATAATTAAGGAGTTGA 350

QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
DB 351 GTGGCCCGTTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTTTC 410

QY 132 a-----AlaValAsnGlyAlaTyrThr 140
DB 411 AAGCCAGAGCTCAGATGCTACTGAGCTCTAAGCTTAAGTCAGGCTGTTAATCGAGCGTGGATGC 470

QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
DB 471 T-----GGATTAGTGTGTGTGTTGCCGC 494

QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
DB 495 TGGAAACAGTGGACCTTCAACAGTATACAATCGTCTCCAGCAGCTGCACGCAAGTAT 554

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
DB 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581

QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
DB 582 TGTTATAACAAGCTTCTCAAGCAGAGGGCCAACTTACACAGCAGCTCTGGGACATCAATGGCAACTCTCTCA 752

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
DB 693 GSGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGCAACTCTCTCA 752

QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
DB 753 CGTAGCTGGTATTGAGCCCTCTTCTCTCAA-----GCACACCC 791

QY 280 oLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
DB 792 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
DB 852 AGATGAAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 890

QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
DB 891 TAAATGATACAAAGGTATAAAC-----TACGATAACTATGCAAAGCTAGTGTTCACCTGG 944

QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
DB 945 ATATGTGGCCAAACAAAGCAGCAGCCAACTCACAGTTCGTTATTAGCGAGGCTTCGTTCCGT 1004

QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
DB 1005 ACTGCCCATATTACTGGACAAATGCCAAT----- 1035

QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385

```

Db 1036 -AGGACCTTGATCTTACCTTCAGATCCCAATGAAACCCAG---GTTGACTACTCTTA 1091
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACGCCCTACTAT-----GGATTCGAAAGGTTGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAAACCACTGATGGAACATGACCAATTAAAGTTGTAAAGCTACAGC-----GGAAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
ART85668
ID ART85668 standard; DNA; 1566 BP.
XX
AC ART85668;
XX
DT 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus furiosus protease coding sequence.
XX
XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS Pyrococcus furiosus; DSM-3638.
XX
PH Key Location/Qualifiers
FT CDS 1..1566
FT FT /cag= a
FT FT /transl_except= (pos: 1282. 1284, aa: Xaa)
FT FT /note= "Xaa= Gly, Val"
XX
PN WO9712823-A1.
XX
XX 19-JUN-1997.
XX
XX 07-NOV-1996; 96WO-JP003253.
XX
XX 12-DEC-1995; 95JP-00323285.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
XX Teunabawa S, Kato I;
XX WPI; 1997-332794/30.
XX P-PSDB; AAW24122.
XX
XX Protease(s) and genes encoding them obtained from Thermococcus and
XX Pyrococcus strains - have extremely high thermal stability and are useful
XX industrially and as research reagents.
XX
XX Claim 7; Page 90-91; 159pp; Japanese.
XX
XX This sequence represents the coding sequence for the protease from
XX Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
XX invention. The proteases of the invention have extremely high thermal
XX stability. The proteases can be used as research reagents, and
XX industrially in the food, drug and chemical industries. (Updated on 17-
XX OCT-2003 to standardise OS field)
XX
XX Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 2.11e-22 Length: 1566
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122

```

```

DB: 2 Gaps: 20
US-09-985-689a-1-copy (1-434) x AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 24 GTCTGAGCTCAAGTTATGGAACCTTACGTTTGGACCTGGATATGATGGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyValIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAAGTA-----ATTGGGTGGTAGATTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACCTCATGTAGTCTTAATAGCAGCTGG 230
QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAGGTTCTAGGTGCGGATGGTTCTGGAACATATCTACTATAATTAAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGCGTTGATACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTTCTGTTTC 410
QY 132 a-----AlaValAsnGlyValaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTCAGCGTCTTAAGTCAGGCTGTTTAATGACGCGTGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
Db 471 T-----GGATTAGTTGTGTGTTGCCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 495 TGGAAACAGTGGACCTTACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAGATATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTTATTAACAGCTTCTCAAGCAGAGGGCCCACTGCAGACGGCAGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGCTCCAGAAATGGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro 260
Db 693 GGGTCAACCAATTAATGACTATTATACAGCAGCTCCTGGGACATCATATGGCACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGACGCCCTCTTGTCCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGCAAAAGATAAAACAGCCCTCATAGAACTGCTATATCGTAAAGCC 851
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTACGGTGC-----GGTAGGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330

```

Db 891 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAAAGCTAGTGTCTACTGG 944
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTTGCCAACAAGGAGCCAACTCACAGTTGGTTATTAGCGGAGCTCGTTCCT 1004
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATTACTGGGACAAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1036 -AGGACCTTGACTTCTTACTCTACAGTCCCAATGGAAACCAG--GTTGACTACTCTTA 1091
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyA-gAsnAsnValGluAsnValPheIleAs 405
 Db 1092 CACGCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTATTA 1127
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1128 CAACCCAACTGATGGAACTAGGACAAATTAAAGTTGTAAGCTACAGC-----GGAAG 1178
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 10

AAT85695
 ID AAT85695 standard; DNA; 1962 BP.

XX AC AAT85695;

XX DT 17-OCT-2003 (revised)

XX DT 20-APR-1998 (first entry)

XX DE Pyrococcus furiosus PFUS protease coding sequence.

XX KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX OS Pyrococcus furiosus; DSM-3638.

XX PN W09721823-A1.

XX PD 19-JUN-1997.

XX PF 07-NOV-1996; 96WO-JP003253.

XX PR 12-DEC-1995; 95JP-00323285.

XX PA (TAKI) TAKARA SHUZO CO LTD.

XX PI Takakura H, Moxishita M, Yamamoto K, Mitta M, Asada K;

XX PI Tsunasaawa S, Kato I;

XX DR WPI; 1997-332794/30.

XX DR P-FSDB; AAW24129.

XX PT Protease(s) and genes encoding them obtained from Thermococcus and
 XX PT Pyrococcus strains - have extremely high thermal stability and are useful
 XX PT industrially and as research reagents.

XX PS Disclosure; Page 123-125; 159pp; Japanese.

XX CC This sequence represents the coding sequence for the protease from
 XX CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 XX CC invention. The proteases of the invention have extremely high thermal
 XX CC stability. The proteases can be used as research reagents, and
 XX CC industrially in the food, drug and chemical industries. (Updated on 17-
 XX CC OCT-2003 to standardise OS field)

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,76e-22 Length: 1962
 Score: 411.50 Matches: 140
 Percent Similarity: 42.55% Conservative: 60
 Best Local Similarity: 29.79% Mismatches: 148
 Query Match: 18.31% Indels: 122
 DB: 2 Gaps: 20

US-09-985-689A-1-COPY (1-434) x AAT85695 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 Db 480 CACAAATAGGAATAATGACACTGGAAATGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAen-- 63
 Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 567 TAGGAGTTATCCATACGATACCATGGACATGGAATGCTCCAGGAGCTTAAGCTGGCGG 626
 QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAATGCTCCAGGAGCTTAAGCTGGCGG 686
 QY 94 sGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 687 AATTAAGGTTCTAGGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyValaGlyIleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGCCCGTTGATACAAAGATAGTACGGAATTAAGCTCATTATCTTCTTCTGTTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTCTAAGTCAGGCTGTTAATGCGCGTGGGATGC 866
 QY 140 rAspSerArgAsnValAlaAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaA 160
 Db 867 T-----GGATTAGTTGTTGTTGCGCG 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
 Db 891 TGGAAACAGTGGACCTAAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGCGCAACTGCAGACGCGAGGCTTAAGCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTCTCTCAGAAACTCGAATAATTGCTGCCAGACAAAGT-----GGAATCTAGAT 1088
 QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI1 260
 Db 1089 GGGTCAACCAATTATGACTATTACAGCAGAGCTCTCTGGACATCAATGGCAACTCTCTCA 1148
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 1149 CGTAGCTGTATTGCGGCCCTCTTGTCTCAA-----GCACACCC 1187
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 1188 GAGCTGGAGCTCCAGACAAAGTAATAAACAGCCCTCTATAGAACTGCTGATATCGTAAAGCC 1247

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 DB 1248 AGATGAATAGCGGATATAGCTACGGTGCA -----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
 DB 1287 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTGTTCACCTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 DB 1341 ATATGTTGCCAACAAAGGCGAGCAACTCACCAGTTGGTTATAGCGGAGTTCGTTGCT 1400
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
 DB 1401 AACTGCCACATTATATCGGACAAATGCCAAT----- 1431
 QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 DB 1432 -ACGGACCTTGATCTTACTCTACGATCCCAATGGAAACCAAG---GTTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 DB 1488 CACCGGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 DB 1524 CAACCCAACTGATGGACATGGACATGAATTAAGTTGTAGCTACAGC-----GGTAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 DB 1575 TGCAAACTATCAAGTAGATGTGTAAGT 1602

RESULT 11

AAX05929
 ID RAX05929 standard; DNA; 1962 BP.

XX AAX05929;
 AC AC
 XX XX
 DT DT
 XX XX

06-MAY-1999 (first entry)

DE Hyperthermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX additive; drug; washing agent; foodstuff; chemical synthesis; ds.

OS Pyrococcus furiosus.

XX W09856926-A1.

XX 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP002465.

XX 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX WPI: 1999-080907/07.

XX P-PSDB; AAM94841.

PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Disclosure; Page 59-60; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores: 2,76e-22 Length: 1962
 Pred No.: 411.50 Matches: 140
 Score: 42.55% Conservative: 60
 Percent Similarity: 29.79% Mismatches: 148
 Best Local Similarity: 18.31% Indels: 22
 Query Match: 2 Gaps: 20

US-09-985-689A-1-COPY (1-434) x AAX05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 DB 420 GTCTGCGAGCTCAAGTTATGGCAACTTACGTTTGGAACTTCGGATATGATGTTCTGGAAT 479
 QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 DB 480 CACAATAGGAATAATTACACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 DB 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTCTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 DB 567 TAGGAGTTATCCATACCATGACATGCATGCACTGGAACATCATAGTTCATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 DB 627 TACTGGAGCAAGTAATGCAAGTACAAAGGGATGCTCCAGGAGCTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 DB 687 AATTAAAGTTCTAGTGCCTGCTGTCGAGCATATCTACTATATAATTAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 DB 747 GTGGGCGCTTGATTAACAAAGATAAGTACGGAATTAAAGTTCATTATCTTCTCTGTTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 DB 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACGGTGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 DB 867 T-----GGATTAGTTGTTGTTGTTGCTGCGC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaAl 180
 DB 891 TGGAAACAGTGGACCTTACAAAGTATACATCGTTCTCCAGCAGCTGCAACCAAGTTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 DB 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
 QY 200 rHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 DB 978 TGTATTAAACAGCTTCTCAACGACAGGGCCCAACTGCAGACGGCGGCTTAAGCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
 DB 1038 TGTGTCTCCGGAACCTGATTAATGCTGCCAGAGCAAGT-----GGAAGTACGAT 1088

QY 240 eTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
 DB 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGCACTCTCA 1148
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLleThrPr 280
 DB 1149 CGTAGCTGGATTGCGACCCCTCTGTCCAA-----GCACACCC 1187
 QY 280 Olys-----ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla----- 293
 DB 1188 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPglyArgVa 310
 DB 1248 AGATGAATACCGCATATACCTACGCTGCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 DB 1287 TAATGCATACAGCTATAAAC-----TAGGATACTATGCAAGCTAGTGTCTACTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 DB 1341 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTTCGTTATTAGCGGAGCTTCGTTCGT 1400
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 DB 1401 AACTGCCACATTATCTGGGACAAATGCAAT----- 1431
 QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 DB 1432 -AGCGACCTTGACTTTTACCTCTACGATCCCAATGGAAACAG---GTTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIleAs 405
 DB 1488 CACCGCTTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 DB 1524 CAACCACTGATGGAACATGGACAAATTAGGTTGTAAAGCTACAGC-----GGAAG 1574
 QY 425 cGlnThrPheSerLeuAlaIleValAsn 434
 DB 1575 TGCAAACTATCAAGTAGATGGTAAAGT 1602
 RESULT 12
 AAT85669
 ID AAT85669 standard; DNA; 1977 BP.
 XX AC AAT85669;
 XX DT 20-APR-1998 (first entry)
 XX DE Protease coding sequence.
 XX KW Protease; research reagent; thermal stability; ss.
 XX OS Synthetic.
 XX PN WO9721823-A1.
 XX PD 19-JUN-1997.
 XX PF 07-NOV-1996; 96WO-JP003253.
 XX PR 12-DEC-1995; 95JP-00323285.
 XX PA (TAKI) TAKARA SHUZO CO LTD.
 XX PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 XX PI Tsurasawa S, Kato I;
 XX DR WPI; 1997-332794/30.
 XX DR P-PSDB; AAM24123.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 11; Page 95-97; 159pp; Japanese.
 PS This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries
 XX Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,78e-21 Length: 1977
 Score: 398.50 Matches: 133
 Percent Similarity: 43.07% Conservative: 69
 Best Local Similarity: 28.36% Mismatches: 147
 Query Match: 17.73% Indels: 120
 DB: 2 Gaps: 20

US-09-985-689A-1-COPY (1-434) x AAT85669 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
 DB 420 GTCTGCAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGATATGATGTTCTTGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
 DB 480 CACAATAGGAATAATTGCACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 DB 522 AGATCTCCAAGAAAGTA-----ATTGGTGGTAGATTTTGTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
 DB 567 TAGGAGTTATCATCAGATGACCATGAGTGGACTCATGTAGTTCATAGCAGCTGG 626
 QY 78 ValGlnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 DB 627 TACTGGACGCAAGTAATGGCAAGTACAGGAATGCTCCAGAGAGTAACTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG 112
 DB 687 AATTAAGGTTCTAGTGGCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
 DB 747 GTGGCCGTTGATACAAAGATAAGTACGAATTAAGTTCATTAATCTTCTTGGTTC 806
 QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
 DB 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGGCCGCTCAACACCGCTGGAGCG 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 DB 867 C-----GGTAGTAGTGTCTGCTGCGCC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaAl 180
 DB 891 CGGCAACAGCGGGCGCAACACCTACACCTCGCTCAGCCGCCGCGCGAGCAAGTCA 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 DB 951 AACCGTCGGTGCA-----GTTGACAGCAACGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 DB 978 CAACATCGCAGCTTCTCCAGCAGGGGACCGACCGGAGGCTCAAGCGGAGCT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240

```

Db      1038  CGTCGCCCGCGGCTTGACATCATAGCCCGCGCCAGC-----GGAACGAGCAT 1088
Qy      240  eTTPAlaAsnHieAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
Db      1089  GGGCACCCCGGATAACGAGCTACTACCAAGGCGCTCTGGAACAGCATGGCCACCCGCA 1148
Qy      260  eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeu 279
Db      1149  CGTTTCGGGCGTGGCGCGCTCATCTCCAGGCCAC----- 1185
Qy      279  rProLys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db      1186  -CGAGCTGACCCCGGACAGGTGAAGACCGCCCTCATCGAGACCGCGGACATAGTCG 1244
Qy      294  -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
Db      1245  CCCCAAGGAGATACCGGACATCGCTACGCTGCG-----GGTAG 1283
Qy      309  gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSe 326
Db      1284  GGTGAAGCTCTACAGGCCATCAAGTACGACGACTACGCCAAGCTCACCTCCACGGCTC 1343
Qy      326  rLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeu 346
Db      1344  CGTCGCCGACAGGGAAGCGCCACCCACACCTTCGACGTCAGCGCGGCCACCTCGTGC 1403
Qy      346  sIleSerLeuValTyrSerAspAlaProLaserThrThrAlaSerValThrLeuValAs 366
Db      1404  CGCCACCTCTACTGGAC----- 1433
Qy      366  nAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe 386
Db      1434  CGACATCGACCTCTACTCTACGACCCCAACGCGGAACGAG--GTTGACTACTCTCTAC 1490
Qy      386  rSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsn 406
Db      1491  CGCCTACTAC-----GGCTTCGAGAAGGTGGCTACTACAA 1526
Qy      406  aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPro 426
Db      1527  CCCGACCGCGGAACCTGGACGTCAGGTCAAGTTCGTCAGCTACAAG-----GGCGCG 1577
Qy      426  nThrPheSerLeuAlaIleValAsn 434
Db      1578  GAATACACGAGTCGACGCTGCTGACG 1602

RESULT 13
ABLS4900
ID      ABL54900 standard; DNA; 2121 BP.
AC      ABL54900;
XX
XX
DT      11-SEP-2003 (revised)
DT      31-MAY-2002 (first entry)
XX
XX
DE      T. yonsei subtilisin-like serine protease coding sequence.
XX
XX
KW      Subtilisin-like serine protease; ss.
XX
XX
OS      Thermoaerobacter yonseiensis.
XX
XX
FH      Key Location/Qualifiers
FT      CDS 142..1779
FT      /*tag= a
FT      /product= "subtilisin-like serine protease"
XX
XX
XX      KR2000072141-A.
XX
XX
XX      05-DEC-2000.
XX
XX      04-AUG-2000; 2000KR-00045411.
XX
XX

```

```

PR      04-AUG-2000; 2000KR-00045411.
XX
XX      (KIMY) KIM Y S.
XX
XX      Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX      WPI; 2001-298092/31.
XX      P-FSDB; ABB09483.
XX
XX      New DNA sequence of thermophilic protein decomposition enzyme and protein
XX      derived therefrom.
XX
XX      Claim 1; Page 6; 15pp; Korean.
XX
XX      This sequence represents the DNA encoding the Thermoaerobacter yonsei
XX      subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX      to standardise OS field)
XX
XX      Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;
XX
Alignment Scores:
Pred. No.:      3,29e-19      Length:      2121
Score:          372.00      Matches:    130
Percent Similarity: 46.29%      Conservative: 51
Best Local Similarity: 33.25%      Mismatches: 114
Query Match:    16.56%      Indels:     96
DB:             4           Gaps:       19

```

US-09-985-689A-1-COPY (1-434) x ABL54900 (1-2121)

```

Qy      6 GlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIle 25
Db      445  GGAATCAAAA-----GCACGAGTGATTTGGAGTCACAGGAAAAAATAAACA 495
Qy      26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
Db      496  ATACAAATTATGACAGGTATAGACGGAATCACGTTGACCTCA----- 543
Qy      46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db      544  -GGTGGAAAAATA-----ATAGATGGAAGACTTTATCAACACAAA 585
Qy      64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db      586  ACTACACATACGACGACAAATGCCATGGAACCTCAGTACAGTATTCGTCGAGGTACA 645
Qy      80 GlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 96
Db      646  GGTGCTGGAACACAGCTTTTACAAAGCGCTGCTCCTGATGCTTTGTTGGTAGGAATAAAA 705
Qy      97 IleMetAspSerGlyGlyGly-----LeuGlyGlyLeuProSerAsn 110
Db      706  GTTTTAGTCAAAATGGAAGCGCAGCATGAGCAGTGAACGAGCAATTCAGTGGGCT 765
Qy      111 LeuGln-----ThrLeuPheSerGlnAla 118
Db      766  GTTCAAAATAAGATGATATACGGAATCAAAGTTATAAATTAAGCTCGGCACCTTCTACA 825
Qy      119 TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
Db      826  AGTTCTCATGGA-----ACTGACTTACTCTCATTAGCAGTGAAT----- 864
Qy      139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
Db      865  -----AGAGCAGTAGAT-----AGCGGTATTGTAGTAGTTGTA 897
Qy      159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db      898  GCACAGAGAAACTCTGGCCCTGCAAAATACACCATAGGGTCCCTCGTGTGCGGAAAAA 957
Qy      179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
Db      958  GCCATAACAGTCGACGCAATGGCAGATGTA-----GGTGAACCTGGCTTTAAC 1005

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 20:31:52 ; Search time 83 Seconds

(without alignments)
2901.791 Million cell updates/sec

Title: US-09-985-689a-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVPGQTSLAIWN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/SWOPE985689/runat_10032004_112807_19395/app_query.fasta_1.583
-DB=Issued Patents NA -CPMT=fastap -SUFTX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=SWOPE985689 @cgn_1_1_44@runat_10032004_112807_19395 -NCFU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents_NA.*
1: /cgn2_6/prodata/2/ina/SA COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCFUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2242	99.8	1923	4	US-09-509-814A-5
2	2237	99.6	1923	4	US-09-509-814A-7
3	2178	96.9	1920	4	US-09-509-814A-3
4	2120.5	94.4	3003	2	US-08-873-479-41
5	447.5	19.9	1977	3	US-08-894-818B-2
6	447.5	19.9	1977	4	US-09-445-472-11
7	411.5	18.3	1236	4	US-09-445-472-2
8	411.5	18.3	1566	3	US-08-894-818B-4
9	411.5	18.3	1962	3	US-08-894-818B-34
10	411.5	18.3	1962	4	US-09-445-472-15
11	398.5	17.7	1977	3	US-08-894-818B-6
12	353	15.7	2539	3	US-09-000-016-3

13	353	15.7	2539	4	US-09-514-340-3	Sequence 3, Appli
14	348	15.5	2809	3	US-09-000-016-1	Sequence 1, Appli
15	348	15.5	2809	4	US-09-514-340-1	Sequence 1, Appli
16	311.5	13.9	2532	1	US-07-671-376C-4	Sequence 4, Appli
17	307	13.7	2835	1	US-08-750-532-2	Sequence 2, Appli
18	307	13.7	4765	1	US-08-750-532-8	Sequence 8, Appli
19	307	13.7	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	307	13.7	4765	4	US-09-445-472-5	Sequence 5, Appli
21	302.5	13.5	1853	3	US-08-894-818B-15	Sequence 15, Appli
22	296	13.2	898	1	US-08-750-532-7	Sequence 7, Appli
23	290.5	12.9	564	1	US-08-750-532-11	Sequence 11, Appli
24	290.5	12.9	564	3	US-08-894-818B-14	Sequence 14, Appli
25	276	12.3	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	276	12.3	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	256.5	11.4	1194	4	US-09-328-352-3407	Sequence 3407, Ap
28	256.5	11.4	10216	2	US-08-875-154-1	Sequence 1, Appli
29	252	11.2	1628	4	US-09-634-238-212	Sequence 212, App
30	251	11.2	840	1	US-08-434-255-7	Sequence 7, Appli
31	251	11.2	840	1	US-08-459-967-7	Sequence 7, Appli
32	251	11.2	840	1	US-08-460-327-7	Sequence 7, Appli
33	251	11.2	840	1	US-08-459-871-7	Sequence 7, Appli
34	251	11.2	840	3	US-09-024-532-1	Sequence 1, Appli
35	251	11.2	840	4	US-09-104-623A-1	Sequence 1, Appli
36	251	11.2	840	4	US-09-019-532-1	Sequence 1, Appli
37	251	11.2	840	4	US-09-417-359A-1	Sequence 1, Appli
38	251	11.2	840	4	US-09-705-185-1	Sequence 1, Appli
39	251	11.2	840	4	US-09-338-746-1	Sequence 1, Appli
40	251	11.2	1110	1	US-08-434-255-5	Sequence 5, Appli
41	251	11.2	1110	1	US-08-459-967-5	Sequence 5, Appli
42	251	11.2	1110	1	US-08-460-327-5	Sequence 5, Appli
43	251	11.2	1110	1	US-08-459-871-5	Sequence 5, Appli
44	251	11.2	1191	1	US-08-434-255-3	Sequence 3, Appli
45	251	11.2	1191	1	US-08-459-967-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-08CT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 2,92e-224 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1

739 TCATGTCATGAAGCCTTCCGCGGGGAAAAATTACTGCATTATATGCAATGATGCGGACGGAGCAT 798
61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
799 AATGCCAATGATACGAATGGTCAATGTCATGCGATGGCTGGCTCCGTATTAGCAAAACGCG 858
81 SerThrAsnGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
859 TCACCTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAACTCATCATGATGATAGC 918
101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySer 120
919 GGTGGGGGACTTCGAGGACTTACCTTCGAATCTGCAAAACCTTATTACGCAAGCATACAGT 978
121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThr 140
979 GCTGGTGCCAGAATTATCAAACTCTCGGGGAGCAGCATGAATGGGGCTTACCAACA 1038
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1039 CATTCGGAATGTGGATGCTATGTGGCAAAATGATATGACGATCCTTTTCGTGCGC 1098
161 GlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
1099 GGGAAATGAGGACCGAAGCGGAAACCATCAGTGCACACGACGACGCTAAAATGCAATA 1158
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACCATATCAAC 1218
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1219 CATGTGGCACAGTCTCTTCCACGTGGACCGACAAAGGATGCGCGATCAAAACCGGATGC 1278
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
1279 ATGGCAACGGGACGTTCACTATCATCATGAGCAAGATCTTCTTCCACCGGATCCTCCTTC 1338
241 TrpAlaAsnHisAspSerLysTyrAlaTyMetGlyThrSerMetAlaThrProIle 260
1339 TGGCGGAACCATGACATGAATAATGATCATATGCGGTGGAACGCTCCATGCGTACACCGATC 1398
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTCTCGGAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGCATCACACCA 1458
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
1459 AAGCTTCTCTATTAAAGCGGCACATGATTCGCGGTGCACATCGGCTTGGCTAC 1518
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
1519 CCGAACGGTAACCAAGATGGGACGAGTGCATTTGGATAAATCCCTGAAACGTTGCTAT 1578
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrThrSerPheThrAlaThr 340
1579 GTGAACAGATCCAGTCTCTATCCACCGCCAAAGACGACGCTACTCTGTTTACTGCTACT 1638
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
1639 GCCGGCAAGCTTTGAAATCTCCCTGATGTGTCTGATGCCCTCGAGCACAACTGCT 1698
361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1699 TCCGTAAACGCTTGCAATGATCTGGACCTTGCTATTAACCTTCAATGGCACACAGTAT 1758
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
1759 GTAGGAATGACTTACTTCGCCATACAATGATAAATGGGATGGCCGCAATAACGTAGAA 1818
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
1819 AATGTATTATTAAATGACCAACAAAGCGGACGCTATACAAATTGAAGTACAGGCTTATAAC 1878

421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1879 GTACCGGTGGACCACAGAACTTCTCGTTGGCAATTGTGAAT 1920
RESULT 3
US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KASEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3
Alignment Scores:
Pred. No.: 1,41e-217 Length: 1920
Score: 2178.00 Matches: 417
Percent Similarity: 99.31% Conservative: 14
Best Local Similarity: 96.08% Mismatches: 3
Query Match: 96.93% Indels: 0
DB: 4 Gaps: 0
US-09-985-689A-1-COPY (1-434) x US-09-509-814A-3 (1-1920)
QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 616 AATGATGTGGCCAGAGGTATTGTCAAAGCGGATGTGGCACAGACGACGTACCGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 676 GGACAAGCCAGATGTGTCGAGTTGCCGATACTCGATTTGGATACAGGAGAAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 736 TCGATCATGATGAGCTTCCTCCGCGTAAATACACACATATATGCTGGGTCCGACCAAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 796 AATGCGAATGATACGAACCGTCAATCCCATGTCGAGTTCCGATTTAGGAAATGGC 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 856 GCAACCAATTAAGGATGGACCTCAAGCGAATCTGTTTTTCAATCCATCATGATGATAGC 915
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySer 120
DB 916 AGTGTGGGCTTGGAGGCTTGCCTTCCAACTTCCAACTTATTCAGCAAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThr 140

976 GCAGGTGCCAGAAATTCATACAAATCTCTGGGGGCGAGCGTGAATGGGGCTCACACGACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAala 160
Db 1036 GATTCAGAAATGTGGTACATATGTAGGAAATATGATGACGATCTCTTTTCGGGCT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GCGAATGAAGGCGGAAAGCGGTACCATCAGTGCACCTGGTACCGCTAAAGAAAGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGGCGAACCGAAACCTGCTCCAGCTTCGGTTCCTATGCGAATATTAAC 1215
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CAGCTTGACAGTCTCTTCCTCGGTGGCGCGACAAAAGATGGCGGAATCAAGCCTGATGC 1275
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGACATACATATTTATCAGCAAGATCTTCTTTCACCCGATTCCTCTTC 1335
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGGAATCATGACAGCAATATGCTATATGGTGGAACTCCATGGCAACCGGAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGGGGAATGTGTCAGCTCGGTGAGCATTTTGTGAAATAATAGAGGAATCACTCCT 1455
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCTATTGAAGCAGCTTGTATGTCAGTGTCTGTGATGTTGGATGGTTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGAAACCAAGATGGGGCGAGTGACCTCGGATAAATCGTTGAACGTTGCCCTAT 1575
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGACGAAATCCAGTCCCTATCACTAGCCAAAAGCGACATATACCTTACTGCAAG 1635
QY 341 AlaGlyProLeuLysLysSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCAATGAAATCTCCCTGTGTGTGGATGGCCCTGCAAGCACTACTGCT 1695
QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAACTGGTCAATGATTTGGATTGGTTCATTCAGCACCAACCGAACACAGATAT 1755
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCCGGAATGACTTCTCAGCACCATTTGACAATACTGGGATGGCGCAATAACGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTTATTTAATTTCGCCCAAGGTGGAACATATACCATTTGAGTGTGCAAGCATATAT 1875
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTCCCGTTGGACCAACAAATCTCTCGTTGGCAATTTGTGAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lytne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agris, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:

Pred. No.: 2 91e-211 Length: 3003
Score: 2120.50 Matches: 405
Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 2 Gaps: 1

US-09-985-689A-1-COPY (1-434) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGTGGCATTTGTGAAGCAGACGTGCGCAAAATACCTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGCAAGERCAGATTGTAGCAGTTGCTGATCTGCTGGCTTGATACAGGAAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGTCATGAAGCATTTCCCGGTAAAGATTACCGCACTATATGCTGGCAGACGAAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACCCCAATGATCCAAATGACATGGACCCCATGTTGCTGGATCTGTGTAGGAAT--- 1706
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATAAAGGATGGCAGCCGCAAGCCAAATCTAGTCTTTCAATCTATTATGGATAGT 1766
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCTGGAGGACCTCTGCTATCTACAAACATTTATTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 1827 GCTGGAGCAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACCGTGCCTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAala 160
Db 1887 GACTCTCGAAATGTTGATGATGATGAGAAATAATGATGACGATCTCTTTTTCGGGCC 1946

161 GlyAsnGluGlyProGlyThrLeuSerAlaProGlyThrAlaLysAsnAla 180
1947 GGAATGAGGACGAGTAGCGTCAATCAGTGCACGAGACCAAAATGCGATT 2006
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnLeu 200
2007 ACAGTTGGGCAACCGAARACCTACGTCGAGTTCCGATCTATCGGATATATAC 2066
201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgLeuLysProAspVal 220
2067 CATGTTGCTCAATCTCTTCCAGAGGTCCTACTAGAGATGGAGTATTAAAGCGGACGTC 2126
221 MetAlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
2127 ATGGCACGAGTACGATATCTCTCTGCTAGATCATCATCTAGCTCCAGATCTCTCATTC 2186
241 TrpAlaAsnHisAspSerTyAlaTyMetGlyThrSerMetAlaThrProIle 260
2187 TGGCAAAACCATGATGATGAATATGCTACATGGTGGTACTTCTATGCTACTCCAAAT 2246
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPro 280
2247 GTACGAGTAAATGTGTCACAAATTAAGGGAGCATTTTGTGAAAATAGAGGGGTAACTCT 2306
281 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAspIleGlyLeuGlyTy 300
2307 AAGCTTCCTTTTAAAGCTGCTTTAAATGCGAGTGTGCGATCTTGACCTTGCTTT 2366
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
2367 CCAATGGTAAACCAAGGATGGGAAGATGAGTAACTGATATAATCCCTAAATGTCGCAAT 2426
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTySerPheThrAlaThr 340
2427 GTGATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATTTCTGTTTACGGTCAA 2486
341 AlaGlyLysProLeuTyIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
2487 GCTGGTAAACCTTTAAATATATCATTGTTGGTTCAGATGCACAGGATGACGACGCGCA 2546
361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTy 380
2547 TCCTAACTTTAGTGAATGATTTAGACTTAGTAACTACTGCAATGGAATGAATCAATAC 2606
381 ValGlyAsnAspPheThrSerProTyAsnAsnTrpAspGlyArgAsnAsnValGlu 400
2607 GTCGAAATGACTTTACAGCACCGTATGATAACAATTTGGGATGGCAGAAACAACGTTGAA 2666
401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
2667 AATGTGTTTATCAATGCTCTCAAGCGGAACGATATACATCGAAGTGCAGGCTTACAAT 2726
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
2727 GTACCAGTAAAGTCGCGCAACCTTTCTTTAGGATTTGATAC 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822

GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA

US-08-894-818B-2

Alignment Scores:

Pred. No.: 8,24e-37 Length: 1977
Score: 447.50 Matches: 137
Percent Similarity: 44.66% Conservative: 68
Best Local Similarity: 29.85% Mismatches: 153
Query Match: 19.92% Indels: 101
DB: 3 Gaps: 18

US-09-985-689A-1-COPY (1-434) x US-08-894-818B-2 (1-1977)

QY 8 VallysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnIleValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGGTACGACGAGCGGTGTGGTGGTCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACCGAAC-----CACCCGATCTGAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GCGAAGGTCATAGGTGCTGACGCGCTCAACGCGAGTCCGCCCTACGATGACCGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAACCCACGTTGCGGGTATCGTCCCGAACCAGCGCGGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGGCGTCCCGCGCGCGAGTCTGTCGCGCTCAAGGTTCTCGGTGCGGACGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTCCAGAACACGACCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGTCATCAACTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834

QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGGCGCTCAACAGCGCTGGAGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTGGCTCGCCCGCGCAACAGCGCGCGCAACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGGCTCACCGCGCGCGAGAGGTCTAATACCGTGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCCAGCTTCTCCAGCAGGGA 1005
QY 210 ProThrLysAspGlyArgLysProAspValMetAlaProGlyThrPheIleLeuSer 229
DB 1006 CCGACCGCGGACGGAAGGTCAAGCGGAAGTCTGTCGCCCGCGGTGGACATCATAGCC 1055
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
DB 1066 CCGCGCGCGCAGC-----GGACCCAGCATGGCCACCCCGATACAGTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly-----AsnValAlaGlnLeu 268
DB 1117 AAGGCTCTGGAACAGCATGGCCACCGCGTTCGGCGGTGGCGCTCATCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
DB 1177 CAGGCGCCAC-----CCGAGCTGGACCCCGCAAGAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
DB 1213 ACCGCCCTCATGAGACCGCGCATAGTCGCCCCCAAGAGATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
DB 1273 GGTGCG-----GGTAGGGTGAACGCTCTACAAGCGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
DB 1312 GAGGACTAGCCCAAGCTACCTTACCGGCTCCGTCGCCGCAAGAGGAGCGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLysIleSerLeuValTrpSerAspAlaPro 355
DB 1372 ACCTTCGAGTCAGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
DB 1423 -----ACGGCTCGAGCGACATCGACCTCTACCTCTACGACCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspThrAspGly 395
DB 1462 AACGGGAACGAG-----GTTGACTACTCTCTACCGCTCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIleGlu 415
DB 1501 -----GGCTTCGAGAGGTGGCTACTACACCCCGCGGAACTGGAGCGTCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1555 GTCGTCAGTACAG-----GGCGGGGGAAGTCACTACAGGTGCGACGTCTGTCAGC 1602

RESULT 6

US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikuoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores: 8,24e-37 Length: 1977
Pred. No.: 447.50 Matches: 137
Score: 44.66% Conservative: 68
Percent Similarity: 29.85% Mismatches: 153
Best Local Similarity: 19.92% Indels: 101
Query Match: 4 Gaps: 18
DB:

US-09-985-689A-1-COPY (1-434) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
DB 433 ATAGGGCGCGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCATCTGAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
DB 535 GGCAGGTCTAGCTAGCTGATACGACGCGCTCAACGGCAGGTCTGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
DB 595 GGACACGGAAACCCACGCTTCGGGTATCGTTGCGGAAACCGGCGAGGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
DB 655 ATAGGCGTCCCGCGCGCGGAGCTCTGTCGGCGTCAAGTCTCGGTCCGCGCGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTGACTGGTCTGTCGAGAACAGGACAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
DB 775 GGATAGGTCATCAACCTCTCCCTCGGTCTCCCGAGGTCTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGGCGCTCAACACGCGCTGGAGCGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTGCGTCGCCCGCGCAACAGCGCGCGCAACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGGCTCACCGCGCGCGGAGAGGTCTAATACCGTGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCCAGCTTCTCCAGCAGGGA 1005
QY 210 ProThrLysAspGlyArgLysProAspValMetAlaProGlyThrPheIleLeuSer 229

```
Db 1006 CCGACCGCGGAGGCTCAAGCGGAGTCTGCGCCCGCGGCTTGACATCATAGCC 1065
Qy 230 AlaArgSerSerLeuAlaProAspSerPheThrAlaAsnHisAspSerLysThrAla 249
Db 1066 CCGCGCGCGCAGC-----GGACACGATGGGCAACCCGATAAACGACTACTACACC 1116
Qy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGGAACCGACGATGGCCCGCGACGTTTCGGCGGTGGCGCGCTCATCTC 1176
Qy 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGGACAAAGGTGAAG 1212
Qy 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCCCTCATCGAGACCGCGACATAGTGGCCCCCAAGGAGATACGGGACATCGCCCTAC 1272
Qy 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyValArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGCG-----GGTAGGTGAAGCTCTACAAAGCCATCAAGTAC 1311
Qy 317 ---AsnValAlaTyrValAsnGlnSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GACGACTACGCCAAGCTCACTTCAACCGGCTCGCTCGCGCACAAAGGGAAGCGCCACCCAC 1371
Qy 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaPro 355
Db 1372 ACCTTCAGCTAGCGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC----- 1422
Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCC 1461
Qy 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGly 395
Db 1462 AACGGGAACAG---GTTGACTACTCTACACCGCCTACTAC----- 1500
Qy 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
Db 1501 -----GGCTTCGAGAGGTGCGCTACTACAACCGACCGCGGAACCTGGACGGTCAAG 1554
Qy 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCAGCTACAAG-----GGCGCGGCACTACACGCTGACGCTGTCAGC 1602

RESULT 7
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2
```

```
Alignment Scores: 2,25e-33 Length: 1236
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 4 Gaps: 20

US-09-985-689a-1-COPY (1-434) x US-09-445-472-2 (1-1236)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTTGAAT 83
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACATAGGAATTAATGACACTGGAATTGAC-----GCTTCTCATCC 125
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAAGAAAGTA-----ATTGGTGGGTAGATTGTCATGG 170
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGAGATTATCCATACGATGACCATGGACATGGAACCTCATGTAGCTCAATAGCAGCTGG 230
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGCAAGTACAAGGAATGCTCCAGGAGCTAAGCTGGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGGTTCTAGTCCGATGTTCTGGAAGCATATCTACTATAATTAAAGGAGTTGA 350
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGGCGGTTGATTAACAAGATAAGTACGGAATTAAAGTTCATTAACTTCTCTGTTTC 410
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGCTAGCTCTAAGTCAAGCTGTTAATCCAGCGGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTCTGTTGTTGCGCG 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGAACACAGTGGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGAAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATAACAAGCTTCTCAAGCAGAGGCGCAACCTGCAGCGCAGGCTTAAGCCTGAGGT 641
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTTCCTCCAGAAACTGGAATATGCTGCCAGAGCAAGT-----GGAACATGAT 692
Qy 240 eTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGAACCTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGATTGCGAGCCCTCTTCTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAATAAACAAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 851
```


Db 642 TGTGTCTCCAGGAATAATCTGCGCAGCAAGT-----GGAACATGACAT 692
Qy 240 eTTPAlaAenHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
Db 693 GGTCAACCAATTAATGACATTAACACAGCAGCTCTGCGACATCAATGGCACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 753 CGTAGCTGTATTGCGACCTCTTGTCTCAA-----GCACACCC 791
Qy 280 olys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAAAGCAAGCCCTCATGAACCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 852 AGATGAATAGCGCATATAGCTTACGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGATACAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCTACCTGG 944
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysAProbe 345
Db 945 ATATGTGTGCCAACAAAGCGCAGCAACTCACCAGTTCTGTTATTAGCGGAGCTTCGTCTGT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGGCCACATTAATCTGCGACCAATCCCAAT----- 1035
Qy 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGCGACTTGAATCTTACCTTACCTACGATCCCAATGGAACCCAG--GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCTACTAT-----GGATTCGAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCAACTGATGGAACATGGAACATTAAGTTGTAAAGCTACAGC-----GGAAG 1178
Qy 425 cGlnThrPheSerLeuAlaIleValAla 434
Db 1179 TGCAACTATCAAGTAGATGCTGTAAGT 1206

RESULT 9
US-08-894-818B-34
Sequence 34, Application US/08994818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 737-3528
TELEFAX: (202) 628-5197
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 4,69e-33 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 3 Gaps: 20
US-09-985-689A-1-COPY (1-434) x US-08-894-818B-34 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGACAGTCAAGTTATGGCAACTTACGTTTGGAACTGGGATATGATGTTCTGAAT 479
Qy 24 nIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGCAATTGAC-----GTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAen-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGTGAATTTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATGACGCTGG 626
Qy 78 yAsnGlySerThrAen-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAGTAATGCAAGTACAGGGAATGCTCCAGAGCTAAGCTGGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGTGCGGATGGTCTTGGAGCATATCTACTATATAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCGTTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTTGTGTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGATGCTAGCTGCTTAACTGAGCTGTTATGACAGCTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTGTGTGTGTTGCGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaAl 180

Db 891 TGAACAGTGGACCTAACAGTATACAAATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
Qy 200 rHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATACAAAGCTTCTCAAGCAGAGGCGCAACTGACAGCGCAGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTCTCTCAGGAACTGATATATGCTGCCAGCAAGT-----GGAAGTATGAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGCGAGCCCTCTGCTCCA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGAGCTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
Qy 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAATAGCCGATATAGCTTACCGTCA-----GGTAGGTT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaThrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGATATACAGCTATTAAC-----TACGTAACATATGCAAGAGTAGTCTCACTGG 1340
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGCGAGCCAACTCACCAGTTGTTATTAGCGGAGCTTCGTTCTG 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProLaserThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCATATATCTGGCAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -ACGACCTTGATCTTTACTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1487
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaThrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGACAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGTAAAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:

Pred. No.: 4,696-33 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 4 Gaps: 20

US-09-985-689A-1-COPY (1-434) x US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCGCAACTTACGTTGGAACTTGGGATATGATGGTTCTCGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethISgl 44
Db 480 CACATAGGANTAAATGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAATCATGATGATTCATATAGCAGCTGG 626
Qy 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAAGCAAGTAATGGCAAGTACAGGGAATGGCTCCAGGAGCTTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGTTCTAGTCCCGATGTTCTGGAGCATATCTACTATATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCGCTTCATACAAAGATAAGTACGGAATTAAGTCAATTAATCTTTCTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCGTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTCTGTTGGTGGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TGGAAACAGTGGACCTAACCAAGTATACATCCGTTCTCCAGCAGCTCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
Qy 200 rHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATACAAAGCTTCTCAAGCAGAGGCGCAACTGACAGCGCAGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGCTCCAGGAACTGGATTAATGCTGCCAGAGCAAGT-----GGAAGTATGAT 1088


```
QY 240 eTpalAAsnHisAspSerlystYrAlatYrMetGlyThrSerMetAlaThrProfil 260
Db 1089 GGGTCAACCAATTAATGACATATTACACAGCAGCTCTCGGACATCAATGGCACTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeThrPr 280
Db 1149 CGTAGCTGTATTGCGAGCCCTCTTGCTCAA-----GCACACCC 1187
QY 280 Olys-----ProSerLeuLeuLysAlaLeuLeuAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspLeuGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAATAGCCGATATAGCCGTGCA-----GTTAGGCT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAC-----TACGATTAAGTAAAGCTAGTGTCACTGG 1340
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGAGCCAACTCACCAGTTTCGTTATTAGCGGAGCTTCGTTCTG 1400
QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATATTACTGGACATGCAAT----- 1431
QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACAG--GTTGACTACTCTTA 1487
QY 385 eThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTTACTAT-----GGATTCGAAAGGTTGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGAATTAAGTTGTAAGCTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602
```

RESULT 11

```
US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozi
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
```

```
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6
```

Alignment Scores:

```
Pred. No.: 1,08e-31 Length: 1977
Score: 398.50 Matches: 133
Percent Similarity: 43.07% Conservative: 69
Best Local Similarity: 28.36% Mismatches: 147
Query Match: 17.73% Indels: 120
DB: 3 Gaps: 20
```

US-09-985-689A-1-COPY (1-434) x US-08-894-818B-6 (1-1977)

```
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 420 GTCTCAGCTAAGTATGCGCAACTTACGTTGGACCTGGATATGATGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGAAAAGTA-----ATTGGTGGGTAGATTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAGTAATGCGCAAGTACAGGGAATGCTCCAGAGCTAAGCTGCGCGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuProSerAsnLeuG 112
Db 687 AATTAAGGTTCTAGGTGCGGATGTTCTGGAAGCATATCTACTATAATTAAGGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgile----- 125
Db 747 GTGGGCGGTGTATAACAAAGATAAGTACGGAATTAAAGTCTTAATCTTCTCTTGGTTC 806
QY 126 -----HisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCGCGTCAACAACGCTGGAGCGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 C-----GGTATAGTAGTCTGCGTCGCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 891 CGGCAACACGCGGCGCGAACACCTACACCGTCCGCTCACCCCGCGCGGAGCAAGGTCTAT 950
```


180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspSerLeuAs 200
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgLeuLysProAspVa 220
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 lMetAlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 eTrpAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProil 260
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeTh 279
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 |||||
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 |||||
503 |||||
504 |||||
505 |||||
506 |||||
507 |||||
508 |||||
509 |||||
510 |||||
511 |||||
512 |||||
513 |||||
514 |||||
515 |||||
516 |||||
517 |||||
518 |||||
519 |||||
520 |||||
521 |||||
522 |||||
523 |||||
524 |||||
525 |||||
526 |||||
527 |||||
528 |||||
529 |||||
530 |||||
531 |||||
532 |||||
533 |||||
534 |||||
535 |||||
536 |||||
537 |||||
538 |||||
539 |||||
540 |||||
541 |||||
542 |||||
543 |||||
544 |||||
545 |||||
546 |||||
547 |||||
548 |||||
549 |||||
550 |||||
551 |||||
552 |||||
553 |||||
554 |||||
555 |||||
556 |||||
557 |||||
558 |||||
559 |||||
560 |||||
561 |||||
562 |||||
563 |||||
564 |||||
565 |||||
566 |||||
567 |||||
568 |||||
569 |||||
570 |||||
571 |||||
572 |||||
573 |||||
574 |||||
575 |||||
576 |||||
577 |||||
578 |||||
579 |||||
580 |||||
581 |||||
582 |||||
583 |||||
584 |||||
585 |||||
586 |||||
587 |||||
588 |||||
589 |||||
590 |||||
591 |||||
592 |||||
593 |||||
594 |||||
595 |||||
596 |||||
597 |||||
598 |||||
599 |||||
600 |||||
601 |||||
602 |||||
603 |||||
604 |||||
605 |||||
606 |||||
607 |||||
608 |||||
609 |||||
610 |||||
611 |||||
612 |||||
613 |||||
614 |||||
615 |||||
616 |||||
617 |||||
618 |||||
619 |||||
620 |||||
621 |||||
622 |||||
623 |||||
624 |||||
625 |||||
626 |||||
627 |||||
628 |||||
629 |||||
630 |||||
631 |||||
632 |||||
633 |||||
634 |||||
635 |||||
636 |||||
637 |||||
638 |||||
639 |||||
640 |||||
641 |||||
642 |||||
643 |||||
644 |||||
645 |||||
646 |||||
647 |||||
648 |||||
649 |||||
650 |||||
651 |||||
652 |||||
653 |||||
654 |||||
655 |||||
656 |||||
657 |||||
658 |||||
659 |||||
660 |||||
661 |||||
662 |||||
663 |||||
664 |||||
665 |||||
666 |||||
667 |||||
668 |||||
669 |||||
670 |||||
671 |||||
672 |||||
673 |||||
674 |||||
675 |||||
676 |||||
677 |||||
678 |||||
679 |||||
680 |||||
681 |||||
682 |||||
683 |||||
684 |||||
685 |||||
686 |||||
687 |||||
688 |||||
689 |||||
690 |||||
691 |||||
692 |||||
693 |||||
694 |||||
695 |||||
696 |||||
697 |||||
698 |||||
699 |||||
700 |||||
701 |||||
702 |||||
703 |||||
704 |||||
705 |||||
706 |||||
707 |||||
708 |||||
709 |||||
710 |||||
711 |||||
712 |||||
713 |||||
714 |||||
715 |||||
716 |||||
717 |||||
718 |||||
719 |||||
720 |||||
721 |||||
722 |||||
723 |||||
724 |||||
725 |||||
726 |||||
727 |||||
728 |||||
729 |||||
730 |||||
731 |||||
732 |||||
733 |||||
734 |||||
735 |||||
736 |||||
737 |||||
738 |||||
739 |||||
740 |||||
741 |||||
742 |||||
743 |||||
744 |||||
745 |||||
746 |||||
747 |||||
748 |||||
749 |||||
750 |||||
751 |||||
752 |||||
753 |||||
754 |||||
755 |||||
756 |||||
757 |||||
758 |||||
759 |||||
760 |||||
761 |||||
762 |||||
763 |||||
764 |||||
765 |||||
766 |||||
767 |||||
768 |||||
769 |||||
770 |||||
771 |||||
772 |||||
773 |||||
774 |||||
775 |||||
776 |||||
777 |||||
778 |||||
779 |||||
780 |||||
781 |||||
782 |||||
783 |||||
784 |||||
785 |||||
786 |||||
787 |||||
788 |||||
789 |||||
790 |||||
791 |||||
792 |||||
793 |||||
794 |||||
795 |||||
796 |||||
797 |||||
798 |||||
799 |||||
800 |||||
801 |||||
802 |||||
803 |||||
804 |||||
805 |||||
806 |||||
807 |||||
808 |||||
809 |||||
810 |||||
811 |||||
812 |||||
813 |||||
814 |||||
815 |||||
816 |||||
817 |||||
818 |||||
819 |||||
820 |||||
821 |||||
822 |||||
823 |||||
824 |||||
825 |||||
826 |||||
827 |||||
828 |||||
829 |||||
830 |||||
831 |||||
832 |||||
833 |||||
834 |||||
835 |||||
836 |||||
837 |||||
838 |||||
839 |||||
840 |||||
841 |||||
842 |||||
843 |||||
844 |||||
845 |||||
846 |||||
847 |||||
848 |||||
849 |||||
850 |||||
851 |||||
852 |||||
853 |||||
854 |||||
855 |||||
856 |||||
857 |||||
858 |||||
859 |||||
860 |||||
861 |||||
862 |||||
863 |||||
864 |||||
865 |||||
866 |||||
867 |||||
868 |||||
869 |||||
870 |||||
871 |||||
872 |||||
873 |||||
874 |||||
875 |||||
876 |||||
877 |||||
878 |||||
879 |||||
880 |||||
881 |||||
882 |||||
883 |||||
884 |||||
885 |||||
886 |||||
887 |||||
888 |||||
889 |||||
890 |||||
891 |||||
892 |||||
893 |||||
894 |||||
895 |||||
896 |||||
897 |||||
898 |||||
899 |||||
900 |||||
901 |||||
902 |||||
903 |||||
904 |||||
905 |||||
906 |||||
907 |||||
908 |||||
909 |||||
910 |||||
911 |||||
912 |||||
913 |||||
914 |||||
915 |||||
916 |||||
917 |||||
918 |||||
919 |||||
920 |||||
921 |||||
922 |||||
923 |||||
924 |||||
925 |||||
926 |||||
927 |||||
928 |||||
929 |||||
930 |||||
931 |||||
932 |||||
933 |||||
934 |||||
935 |||||
936 |||||
937 |||||
938 |||||
939 |||||
940 |||||
941 |||||
942 |||||
943 |||||
944 |||||
945 |||||
946 |||||
947 |||||
948 |||||
949 |||||
950 |||||
951 |||||
952 |||||
953 |||||
954 |||||
955 |||||
956 |||||
957 |||||
958 |||||
959 |||||
960 |||||
961 |||||
962 |||||
963 |||||
964 |||||
965 |||||
966 |||||
967 |||||
968 |||||
969 |||||
970 |||||
971 |||||
972 |||||
973 |||||
974 |||||
975 |||||
976 |||||
977 |||||
978 |||||
979 |||||
980 |||||
981 |||||
982 |||||
983 |||||
984 |||||
985 |||||
986 |||||
987 |||||
988 |||||
989 |||||
990 |||||
991 |||||
992 |||||
993 |||||
994 |||||
995 |||||
996 |||||
997 |||||
998 |||||
999 |||||
1000 |||||

RESULT 12

US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338..2539
IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores:
Pred. No.: 9,1e-27 Length: 2539
Score: 353.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.71% Indels: 90
DB: 3 Gaps: 16
US-09-985-689A-1-COPY (1-434) x US-09-000-016-3 (1-2539)
QY 2 AspValAlaArgGlyLleValLysAlaAspValAlaGlnSerSerThrGlyLeuThrGly 21
Db 953 GACAGTCCGTCGCGGAGATCGCGCCCAAGCGGTGGTCCGCC---GGTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGTGAAGATCGCGTCTCGACACCGGTGTCGACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAGGGCGCGGTACCGCGTCCAGAGACTTCACCGCGCGCCCGGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGGCGCAAGGTGGGCGCACCGCACCGTCCGCTCGATCGCGGCGGCGGCGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAGGGCAAGTACAGGGGTGCGCACCCCGCGCGCGGTCCCTCAACGCGCAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnInThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTGCGGACGACTCCGGCATCTCCGCGGCGATGGAGTGG 1285
QY 118 AlaThrSerAlaGlyAlaArgIleHisThrAsnSerThrGlyAlaAlaValAsnGlyAla 137
Db 118 |||||
119 |||||
120 |||||
121 |||||
122 |||||
123 |||||
124 |||||
125 |||||
126 |||||
127 |||||
128 |||||
129 |||||
130 |||||
131 |||||
132 |||||
133 |||||
134 |||||
135 |||||
136 |||||
137 |||||
138 |||||
139 |||||
140 |||||
141 |||||
142 |||||
143 |||||
144 |||||
145 |||||
146 |||||
147 |||||
148 |||||
149 |||||
150 |||||
151 |||||
152 |||||
153 |||||
154 |||||
155 |||||
156 |||||
157 |||||
158 |||||
159 |||||
160 |||||
161 |||||
162 |||||
163 |||||
164 |||||
165 |||||
166 |||||
167 |||||
168 |||||
169 |||||
170 |||||
171 |||||
172 |||||
173 |||||
174 |||||
175 |||||
176 |||||
177 |||||
178 |||||
179 |||||
180 |||||
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 |||||
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 |||||
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 |||||
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 |||||
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||<


```

QY 98 MetaspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCGGTT-----TTGGCGACACTCCGGCATCTCCGGCGGATGGATGG 1285
QY 118 AlatySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
Db 1286 GCGGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACCGGAGACGACGCGGCTGGAGCGGCGGCTGCAAGCTGTCCGCGGAGAGGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTGTTCGCCATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GCGAGCGGCGAGCGGCGGCTTCCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerA---GlyPro---ThrIys 212
Db 1481 -----GTCAGGACAGGACGAGCTCGCGACTTCTCTCCACGCGGCGCGGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGGCATCAAGCGGACGTACCGCTCCGCGGTGGACATCACGCGCGCGCTCGGCG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGCAACGACATCGCGGAGGAGTGGTGAGGAGCGCGGCGGCTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTGATGCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACTGGACTCCGCGGACTGAGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPglyArg 309
Db 1748 CTCACGCGTCCACCAAG---GGCGGCAAGTACACCCGTTTCGAGCAGGGTTCGGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGCAAGCGCTCCAGCAGACCGGTGATCGCGGACCGCGGCTCGGTGAGCTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGGCTCCAGCAGTGGCGGCGCACACCGACGAGCGCGTCAACGACGCTGACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAsp----- 353
Db 1925 AACCTGGCACCGACGACGTCGCTGAGCTGACGTGACCGCGGCGGCGGCGGCGGCGGCG 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGGCGCGCGCGCGGCTTCTTCACGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2044
QY 358 ThrThrAlaSerValThrLeuValAsnAspLeuAsnLeu----- 370
Db 2045 GGCAGCGCTCGCTGACATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGGCTAGCTGGTTCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2146

```

RESULT 14

US-09-000-016-1
 ; Sequence 1, Application US/09000016
 ; Patent No. 6143541
 ; GENERAL INFORMATION:

APPLICANT: AKIRA ARISAWA et al.
 TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
 TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 STREET: 2033 K Street, N.W., #800
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/000,016
 FILING DATE: January 30, 1998
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Warren M. Cheek, Jr.
 REGISTRATION NUMBER: 33,367
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-721-8200
 TELEFAX: 202-721-8250
 TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2809 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 ORIGINAL SOURCE:
 STRAIN: A-914
 ORGANISM: Streptomyces viridosporus
 ORGANISM: Streptomyces antibioticus
 STRAIN:
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 338...2539
 IDENTIFICATION METHOD: E
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2540...2809
 IDENTIFICATION METHOD: P
 US-09-000-016-1

Alignment Scores:
 Pred. No.: 3,56e-26 Length: 2809
 Score: 348.00 Matches: 129
 Percent Similarity: 42.00% Conservative: 52
 Best Local Similarity: 29.93% Mismatches: 167
 Query Match: 15.49% Indels: 84
 DB: 3 Gaps: 13

US-09-985-689A-1-COPY (1-434) x US-09-000-016-1 (1-2809)

QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
 Db 953 GACACGTCGTCGCGGACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1009
 QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
 Db 1010 AAGGCGGTGAAGATCGCGGCTCTGGACACCGGCTCGACACGAGC----- 1054

QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGCGCCGGGTGACCGCTCCACAGAACTTCACCCCGCCGCGCCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGCGACAAAGTGGGCCACGACCCACGCTCCATCGCGCGCGGACGCGGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAAAGCGCTCGCACCGCGCGCGCATCTCAACGCGCAAGTTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCCGT-----TTCCGGCGAGACTCCGGCATCTCCCGCGCATGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGCGCGCGAGCGCGCGCGCTGCTCAACATGAGCTGGCGGCGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCCGCTGGAGGCGCG-GTCCGACAAAGCTGTCCGCGGAGAGGGCGTCTGTTCGC 1404
QY 157 upheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGGCGCGGCA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCCCTCACCTCCGCGCC-----GTCCGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerArgGlyPro---ThrIleAspGlyArgIle 216
Db 1486 CGACAAGGACAACTCGCGGACTCTCTCCACCGGCGCGCGCTCGCGCGCGCGCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTCACCGCTCCCGCGCGTGACATCACCGCGCGCTCGCGGAGGCGCAACGA 1605
QY 236 aAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGGTGCGTAGGACCGCGCGGTATACATCATCTCCGCGCGCTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnAr 276
Db 1666 GCGCACCCCGCACGTCGCGCGCGCGCGCGCTCTGAAGCAGCAG-----1711
QY 276 gGlyIleThrProLysProSerLeuLeuLeuAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGACCTCCCGCGAACTGAAGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG--GGCGCAAGTACACCCCGTTCGAGCAGGGTTCGGCGCGGATCCAGGCGCA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer-----328
Db 1816 CAAGCGCTCCAGCAGACCGTGTATCGCGACCGCGCTCTCGGTGAGCTTCGCGGTCCAGCA 1875
QY 329 -----ThrSerGlnIleAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGAGCGCGGTACCAAGCAGCTGACTACCGCAACCTCGGCAC 1935
QY 340 tAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----353
Db 1936 CCAGGACGTCAGCTGAAGCTGACGTCCAGCCCGCCACCGCCCAAGGCGAGGCGGCC 1995
QY 354 -----AlaProAlaSerThrThrAlaLe 361
Db 1996 GCGCGGCTTCTTCAGCTGGCGCGCACACCGGTGACCGTCCCGCGGCGCGCGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAsnLeu-----Va 371

Db 2056 CGTCGACATACCCCGACACCGCGCTCGCGCGCAGGTGAGCGCGGTACTCGGCGTA 2115
QY 371 lileThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGCGCACGCGCGCGCGCGAGACGTC 2146
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 3,56E-26 Length: 2809
Score: 348.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.49% Indels: 84
DB: 4 Gaps: 13

US-09-985-689A-1-COPY (1-434) x US-09-514-340-1 (1-2809)

QY 2 AspValAlaArgGlyIleValValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGTCCGTCGGCAGATCGGCGCCGCCCAAGCGTGGTCCGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGTGAAGATCGCGTCTCGTACACCGCGTGTTCGACAGAC- 1054
QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAGAGCGGGGTGACCGGTCCTCAAGAACTTCCACCGCGCCCGCCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGCGACAAAGTGGCGCACCGGACCCACCGTCTCGATCGCGCGGCGCACGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAAAGGCGTGGACCCCGCGCGGCGATCTCTCAACGGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGCGACGACTCCGGCATCTCCCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGCGCGCGCGTGTCAATGACGCTGGCGCGGATGGACACACCGGAG 1345
QY 137 atyThrThrAspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLe 157
Db 1346 ACCGACCCGCTGGAGSGCGCG-GTCGACAAAGTGTCCGCGGAGAGGCGTCTCTGTCGC 1404
QY 157 upheAlaAlaGlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGCAACGAGGGCGCGGAG-----TCGATCGGTTCGCCGCGCGCGCA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCACGTCGGCGCC-----GTGGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 CGACAAGGACACAGCTCGCGACTTCTCTCCACCGCGCCCGCTCGCGGAGCGCGCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCCGAGCTCACGCTCCCGCGTGGACATCACGCGCGCTCGCGGAGCGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyThrSerMe 256
Db 1606 CATCGGCGAGGAGTGGTGGAGGACCGCGCGCTACATGACCATCTCCGCGCACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnAr 276
Db 1666 GCGGACCCCGACGTCGCGGCGCGCGCTCTCTGAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuIysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGACTGGACCTCCGCGCGAACTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGGTTCGGCGCGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGlnUserSerSerLeuser----- 328
Db 1816 CAAGCGCTCCAGCAGACCGGTGATCGCGCGACCCCGGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnIysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGCGCACCGCAGCGCGGTGTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935

340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
1936 CCAGGACGTCACGCTGAAGCTGACGTCCAGCCGACCCCAAGGCAAGCGGCGCC 1995
354 -----AlaProAlaSerThrThrAlaSe 361
1996 GCGGGCTTCTTCACGCTGGGCGCCACCCAGGTGACCGTCCGCGCGGCGGCGGCGCTC 2055
361 rValThrLeuValAsnAspLeuAsnLeu-----Va 371
2056 CGTCGACATGACCGCGCGACCCCGCTCGGCGCGGCGGCGGCGGCGGCTC 2146
371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
2116 CGTGTGCGCACGGGCGGCGGCGGCGGCGGCGGCGGCGGCTACTCGGCGTA 2115

Search completed: March 15, 2004, 22:29:59
Job time : 114 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 21:48:38 ; Search time 360 Seconds
(without alignments)
4438.371 Million cell updates/sec

Title: US-09-985-689a-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSGLY.....EVQAVNPVGPQTRSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2432557 seqs, 184079884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USFIO.spool/SWORE985689/runat_10032004_112808_19481/app_query.fasta_1.583
-DB=Published Applications NA -QWFI=fastap -SUFFIX=rpnp -MINMATCH=0.1
-LOOPCL=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=SWOP985689 @CGN 1.1.164 @runat_10032004_112808_19481
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1:	/cgn2_6/ptodata/1/pubpna/PCT_PUBCOMB.seq:
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
19:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2242	99.8	1305	15	US-10-385-662-1	Sequence 1, Appli
2	447.5	19.9	1977	13	US-10-090-624-11	Sequence 11, Appl
3	411.5	18.3	1236	13	US-10-090-624-2	Sequence 2, Appli
4	411.5	18.3	1236	13	US-10-090-624-15	Sequence 15, Appl
5	363.5	16.2	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	363.5	16.2	3624	14	US-10-156-761-1	Sequence 1, Appli
7	348	15.5	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	348	15.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	336	15.0	135638	14	US-10-314-657-1	Sequence 1, Appli
10	332.5	14.8	1329	9	US-09-974-300-1934	Sequence 1934, Ap
11	307	13.7	4765	13	US-10-090-624-5	Sequence 5, Appli
12	306.5	13.6	1560	15	US-10-084-846A-113	Sequence 113, App
13	306.5	13.6	59816	15	US-10-084-846A-2	Sequence 2, Appli
14	306.5	13.6	59816	15	US-10-084-846A-2	Sequence 2, Appli
15	282	12.6	3788	10	US-09-927-827-33	Sequence 33, Appl
16	278	12.4	3743	10	US-09-927-827-29	Sequence 29, Appl
17	276	12.3	1306	9	US-09-966-921A-1	Sequence 1, Appli
18	276	12.3	1330	9	US-09-966-921A-5	Sequence 5, Appli
19	275	12.2	3303	14	US-10-156-761-5384	Sequence 5384, Ap
20	268	11.9	2192	12	US-10-424-599-112429	Sequence 112429,
21	256	11.4	1485	9	US-09-974-300-1938	Sequence 1938, Ap
22	253	11.3	1971	9	US-09-974-300-1935	Sequence 1935, Ap
23	251	11.2	840	14	US-10-209-812-1	Sequence 1, Appli
24	247.5	11.0	2166	12	US-10-344-231-17	Sequence 17, Appl
25	247.5	11.0	2166	12	US-10-363-332A-17	Sequence 17, Appl
26	242.5	10.8	3884	10	US-09-927-827-34	Sequence 34, Appl
27	240	10.7	1140	8	US-08-322-678-11	Sequence 11, Appl
28	240	10.7	1140	16	US-10-323-324-11	Sequence 11, Appl
29	240	10.7	1140	16	US-10-323-324-12	Sequence 12, Appl
30	240	10.7	1143	14	US-10-313-853-6	Sequence 6, Appli
31	240	10.7	2588	12	US-10-344-231-20	Sequence 20, Appl
32	240	10.7	2588	12	US-10-363-332A-20	Sequence 20, Appl
33	240	10.7	1497	8	US-08-322-678-6	Sequence 6, Appli
34	237.5	10.6	1497	9	US-09-060-854B-1	Sequence 1, Appli
35	237.5	10.6	1497	14	US-10-033-325-1	Sequence 1, Appli
36	237.5	10.6	1497	14	US-10-228-572-1	Sequence 1, Appli
37	237.5	10.6	1497	15	US-10-423-649-1	Sequence 1, Appli
38	237.5	10.6	1497	15	US-10-323-324-6	Sequence 6, Appli
39	237.5	10.6	1149	16	US-10-146-905A-7	Sequence 7, Appli
40	235.5	10.5	1494	14	US-10-104-693-1	Sequence 1, Appli
41	235.5	10.5	1494	14	US-09-927-827-30	Sequence 30, Appl
42	233	10.4	3452	10	US-09-927-827-30	Sequence 30, Appl
43	231.5	10.3	1332	14	US-10-156-761-5689	Sequence 5689, Ap
44	230.5	10.3	1140	15	US-10-146-905A-9	Sequence 9, Appli
45	227.5	10.1	1140	9	US-09-920-118-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US200400002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

; PRIOR FILING DATE: 2002-10-18
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 1305
 ; TYPE: DNA
 ; ORGANISM: Bacillus sp. KSM-KP43
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1305)
 ; OTHER INFORMATION:
 US-10-385-662-1

Alignment Scores:
 Pred. No.: 5,696-241 Length: 1305
 Score: 2242.00 Matches: 433
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.78% Indels: 0
 DB: 15 Gaps: 0

US-09-985-689A-1-COPY (1-434) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTTGCGGTGGAAATGCTCAAGCGGATGGCTCAGAGCAGCTACGGGTGTAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
 Db 61 GGCAAGGACRATCGTAGCGGTGCCGATACAGGCTGTGACAGGTGGCAATGACAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValThrAsn 60
 Db 121 TCGATCATGAAGCTTCGCGGGGAAATTAATGCTATTATGCTATGCTGGGACGAGGAAT 180
 QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCCAATGATACGAATGCTCATGGTACGCGATGGCTGGCTCGTATTAGGAAACGCG 240
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 241 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCAATCATCATGATAGC 300
 QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 301 GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAGAACCTTATTCAGCCCAAGCATACAGT 360
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyValaTyrThr 140
 Db 361 GCTGGTCCAGAAATTCATCAAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACA 420
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 421 GATTCAGAAATGTGATGACTATGTGCGCAAAATATGATGACGATCTCTTCGCTGCC 480
 QY 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 481 GGGAAATGAGGACCGGACCGGAAACCATCAGTGCCACAGGACAGCTTAAATCAATA 540
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAATATCAAC 600
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 601 CATGTGGCAGATCTCTTTCACGTGGACCGGACCAAGGATGGGCGGATCAACCGGATGTC 660
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 Db 661 ATGGACCGGAGACGTTCATCTATCAGCAAGATCTTCTTCCACCGGATTCCTTCCTTC 720
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260

Db 721 TGGCGAACCATGACAGTAATAATGCATACATGGGTGGACGTCCTCATGGCTACACCGATC 780
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 781 GTTGCTGGAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGATCACACA 840
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTGCGGTGCGAGCTGACATCGGCTTGGCTAC 900
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGGATAATCCCTGAACGTTGCCTAT 960
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 961 GTGACGAGCTCAGTTCCTCTATCCACGACCCAAAAGGAGCGTACTCGTTTACTGCTACT 1020
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
 Db 1021 GCCGGCAAGCCTTGAATAATCTCCCTGGTATGCTGATGCCCTGCGGACCAACTGCT 1080
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1081 TCGTAACGCTGTCTAATGATCTGGACCTTGCTATACCGCTCCAAATGGCACACAGTAT 1140
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
 Db 1141 GTAGGAATGACTTTACTTCGCCATACATAATACTGGGATGGCGCAATAACGTAGAA 1200
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1201 AATGTATTATTAAATCACCACCAAGCGGAGCTATACAAATTGAGGTACAGGCTTATAAC 1260
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1261 GTACCGGTTGGACACAGACCTTCCTGTTGGCAATTGTGAAT 1302

RESULT 2

US-10-090-624-11
 ; Sequence 11, Application US/10090624
 ; Publication No. US20020132335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMODO, Tomoko
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA=6
 ; CURRENT APPLICATION NUMBER: US/10/090,624
 ; CURRENT FILING DATE: 2002-03-06
 ; PRIOR APPLICATION NUMBER: 09/445,472
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 1977
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-090-624-11

Alignment Scores:
 Pred. No.: 1,4e-39 Length: 1977
 Score: 447.50 Matches: 137
 Percent Similarity: 44.66% Conservative: 68
 Best Local Similarity: 29.85% Mismatches: 153
 Query Match: 19.92% Indels: 101
 DB: 18 Gaps: 18

US-09-985-689A-1-COPY (1-434) x US-10-090-624-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAla 27
Db 433 ATAGGGCGGATACCGCTCTGGAACTCCCTCGGTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACGGGAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGTCTATAGGCTGGTACGACGCGGTCAACGGCAGGTTCGACCCCTACGATACAC 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACAGGAACCCACGTTGCGGTATCGTTGCCGGAAACCGGACGCTTACCTCCAGTAC 554
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGGCGTCCCGCGCGGAAGTCTGTCGGGTCAAGGTTCTCGGTCCGACGGTTGC 714
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCCACCATCATCGCGGGTGTGACTGGGTCTCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAGGTCTACCACTCTCCCTCGGTCTCCAGAGTCCCGACGACCGACTCC 834
QY 130 TrpGlyAlaValAlaAsnGlyAlaTyrThrThrAspSerThrAsnValAspTyrVal 149
Db 835 CTCAGTCAGGCGGTCAACACGCTGGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTCTGCGTCCGCGCGGACAGCGGGCGGACACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTCGGCTCACCGCGCGCGGAGCAAGGTCAATACCGTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACGACACGACACATCCGACGCTTCTCCAGCAGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1006 CCGACCGGGACGGAAGGCTCAAGCGGAAGTCTGCGCGCGGTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGCCAGC-----GGAACGACGATGGGACCCCGGATAAAGCACTACTACCC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGGAACGACGATGCGCACCCCGCACGTTTCGGCGGTGGCGGCTCATCCTC 1176
QY 269 ArgGlnHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCCCAC-----CCGAGCTGGACCCCGGACAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCTCTATGAGACCGCGGACATAGTCCGCCCAAGGAGATAGCGGACATCGCCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGG-----GGTAGGTGAACGTCTACAAGGCGCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335

Db 1312 GACGACTACGCCAAGCTCACCTTACCGGCTCGCTCGCGACAAAGGAAAGCGCCACCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCACGTCAGCGCGCCACCTTCGTACCGCCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAlaAspLeuAsnLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGGACATCGACCTCTACCTCTACGACCCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
Db 1462 AACGGGAACGAG-----GTTGACTACTCTCTACACCGCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
Db 1501 -----GCTTCGAGAAGTCTGCTACTACAACCCGACCGCGAACCTCGAGGTCAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCTAGCTACAAG-----GGCGGGCGAACTACCGAGTCGACGTCGTGAGC 1602

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 5,81e-36 Length: 1236
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 25.73% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 13 Gaps: 20

US-09-985-689A-1-COPY (1-434) x US-10-090-624-2 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 24 GTCGACGCTCAAGTATGCGCACTTAGTTGGGACTTGGGATATGATGTTCTGGAAT 83
QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGlu 44
Db 84 CACAATAGGAATAAATGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCCAAGAAAGTA-----ATTGGGTGGTAGATTTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlu 78

Db 171 TAGGAGTATCCATACGATGACCATGGACATGGAACTCACTAGCTTCAATAGCAGCTGG 230
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaLeuValPh 94
Db 231 TACTGGAGACCAAGTAATGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGGTTCTAGTGGCGGATGGTCTGGAGCATATCTACTATAATTAAGGAGTTGA 350
Qy 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCCGCTGTATACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTCTGCTTC 410
Qy 132 a-----AlaValAsnGlyAlaTyThrTh 140
Db 411 AAGCCAGAGCTCAGATGTAATGACGCTCTAAGTCAGGCTGTTAATCAGGCTGGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTGTGGTGGCGC 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGGAAACAGTGGACCTTCAAGTATACAAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACATTTGGAGCC-----GTTGACAAGTATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCCTGAGGT 641
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAAGTGGTAATGTGCTGCCAGACCAAGT-----GGAACTAGCAT 692
Qy 240 eTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProI1 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGTATGTGAGCCCTCTTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGAGCTCCAGACAAAGTAATAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCCCTAGCGTCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerLeuSerThrSe 330
Db 891 TANTGCATACAGCTTATAAC-----TACGATAACTATGCAAGCTAGTGTCACTGG 944
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCCATATATCTGGGACATGCCAAT----- 1035
Qy 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1036 -AGCAGCTTGATCTTTACCTCTACGATCCCAATGCCAATGCCAAGCAAGCAAGCAAGCAAG 1091
Qy 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405

Db 1092 CACCGCTACTAT-----GGATTCCGAAAGGTTGGTTATTA 1127
Qy 405 nLalaProGlnSerGlyThrTyThrIleGluValGlnAlaTyRhsenValProValGlyPr 425
Db 1128 CAACCCAACTGATGAACATGACAAATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TCCAAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 4
US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1993-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15

Alignment Scores:
Pred. No.: 1,17e-35 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 13 Gaps: 20

US-09-985-689A-1-COPY (1-434) x US-10-090-624-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyG1 24
Db 420 GTCTGACGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethIleG1 44
Db 480 CACATAGGATTAATGACACTGGATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTCATCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGAGATTATCCATACCATGACCATGGACATGGAATCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGTGGCTCCAGTCTTCTGGAAGCATATCTACTATATTAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132

[illegible]

```

QY 285 LeuLeuAlaAla-----LeuLeuAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
Db 6919614 ATGTCACCGTCCGAGCAACTCGACGCTCCGTATATACAGTTGCGGCGCGGTCTCAGT 6919673
QY 303 GlyAsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
Db 6919674 GTCCCGAGCGCGTCCGCGCGCGTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919733
QY 321 -----ValAsnGluSerSerSerLeuSerThrSerGln 331
Db 6919734 CGCTGCGCCCATGACGCGCATGACCGCTCACGAGACGCTCACCTACTCCAACTCTCTCC 6919793
QY 332 LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 6919794 GACACACGCGTGCAGTTGAGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919853
QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
Db 6919854 GCGGACACCGCACTCACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 6919913

RESULT 7
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 3,5e-28 Length: 3417
Score: 348.00 Matches: 139
Percent Similarity: 44.75% Conservative: 57
Best Local Similarity: 31.74% Mismatches: 170
Query Match: 15.49% Indels: 74
DB: 14 Gaps: 17

US-09-985-689A-1-COPY (1-434) x US-10-156-761-3306 (1-3417)
QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 745 GGGTACGACGCAAGGCGGTCAAGATCGCGTCTCTGGACACCGGTGTCGAC----- 795
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 796 -----GGGACCCACCGGACCTCAAGGACCGAGGTGGCGGCGGAGTCCAAAGAACTTCTCC 846
QY 58 ArgThrAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 847 CCGCGCGCGGCGCGCGCGCGCACTTCGTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 906
QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93

```

```

Db 907 GGCACCGCGCCCAAGTCCAAAGCGCAAGTCAAGGGTGTCCGCGCGGCGGCGGCGGCGGCGGCTCCTC 966
QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
Db 967 AACGGCAAGGTCTCGACGACACCGGC-----TCCGCGACGACTCCGCGCATCTCGGCC 1020
QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
Db 1021 GGCATGGAGTGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080
QY 133 aValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
Db 1081 GACACCGCGGAGATCGACCGCTCGAAGCGGAG-GTCAACAAGCTCTCCGAGGAGAAGG 1139
QY 153 pMetThrIleLeuPheAlaGlyAsnGlyLysProAsnGly---GlyThrIleSerAl 172
Db 1140 CATCTCTTCGCGATCCGCGCGCGGCAACGAGGCGGAGTTCGCGGAGCGAGACCATCGGCTC 1199
QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 1200 CCGGGGCGGCGCGCGCGCGCTCACCGTCCGCGCC----- 1236
QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 1237 -----GTGAACGACGACGCGCATCGACAGAGTCCGCGCGGCGGCGGCGGCGGCGGCT 1286
QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1287 GCGCGGCGCATCAAGCCCGACGTCACCGGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1341
QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
Db 1342 ----GCGCGCGCGCGCGCGCTCATCGACAGAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCTACCT 1397
QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
Db 1398 CACCATCTCCGATAGTCGATGGGCGACCCCGCATGTCCGCGGCGGCGGCGGCGGCGGCGGCGG 1457
QY 269 gGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
Db 1458 GCAGCAGCAC-----CCCAACTGGTGTCTCGCGGAGCTCAAGGG 1496
QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
Db 1497 CGCGCTGACCGGCTCCGCGAAG-----GGCGGCAAGTACACGCGCTTCAGCGGCGGCTC 1550
QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLe 327
Db 1551 GGGCGGTATCGCGGTCCGACGAGCGATCAAGCAGTCCGTGATCGCAACCCCGCAACTCGGT 1610
QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
Db 1611 G-----AGCTTCGGATCCAGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1646
QY 347 leSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
Db 1647 CAAGCGGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1700
QY 367 sPLeuAsnLeuVal-----IleThrAlaProAsnGlyThrG 379
Db 1701 G-CTGAACCTCCGCTCGACGCGGCGCAACCCCAAGGCGGTCCGCGCTCCGCGGCGCTCT 1759
QY 379 lnTyrValGly---AsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnA 398
Db 1760 TCAAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1807
QY 398 snValGlu-----AsnValPheIleAsnAlaProGlnSerGlyThrThrIleG 415
Db 1808 CGGTCCGATTCACGCTCAACAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1867
QY 415 luValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431

```


DB:	9	Gaps:	13
US-09-985-689A-1-COPY (1-434) x US-09-974-300-1934 (1-1329)			
Qy	11	AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAsp	30
Db	406	GAAGTGTCTAGAAACAATCAGACGCTCAGAGGCAAGGAGTGACATCGCTGTCATTGAT	465
Qy	31	ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle	50
Db	466	ACGGGCGTA-----TACCCTCACCAGAGATCTTGAAGCGCAGGATC	504
Qy	51	ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGly	69
Db	505	AGGCGCTTTTCAAGACTTTATCAACAGCAGACAGAACCTTATGATGACATGGCAGGC	564
Qy	70	ThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn-----LysGly	85
Db	565	ACACACTGCGCGGTGATGTTGCGGAAACGAGCGGCCCATCGGTGAGTACCGCGGA	624
Qy	86	MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly	105
Db	625	CCTCCTCCTGAAGCAGAACTTGTGGTGTAAAAAGTATTGGACAAATGGGATCC---GGA	681
Qy	106	GlyLeuProSerAsnLeuGln-----SerTrpGlyAlaAlaValAsnGly	136
Db	582	TCGCTCGAACCGTCATTCAGGCGTAGATTGGTGATTCAATCAATAGGAAATCCT	741
Qy	113	-----ThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn	128
Db	742	GATGATCCGATCGACATATTTCATTCATTTGGGTGCAGAAGCCTTGGCGCTACGAGAAT	801
Qy	129	-----SerTrpGlyAlaAlaValAsnGly-----	136
Db	802	GAAGAAGAGATCCAGTCGTTAAAGCTGTTCAATCCAGCATGGGACGCGGCAT---	855
Qy	137	AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle	156
Db	856	-----GTTGTA	861
Qy	157	LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla	176
Db	862	TGTCGGGACGCGGCAACTCCGGTCTCTGATGCGCAACAGATTGCCACCGGGTGTGCG	921
Qy	177	LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAla	196
Db	922	ACCAAGATTATTACAGTCGGAGCCTTGGATGAC---AGGGATACAGTCACCCGGGAGGAT	978
Qy	197	AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle	216
Db	979	GACGAT-----GTGCGCTCTATTTCAGCAGAGCGCGCAATCTATGTGTCAAGTC	1029
Qy	217	LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer-----SerLeu	234
Db	1030	AAACCGGACTTGCTGGTACCGGCAAAAATATTACGTCTCGTTCGTTACCCGGATCTTTT	1089
Qy	235	AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr	254
Db	1090	CTCGATAAGCTGCAAAAAACAAACAGAGTCGGCAAAAAATATATGACATTTGTCGGGAACC	1149
Qy	255	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys	274
Db	1150	TCGATGGCTACCGCATCTCGCAGGAATTCGCCCA-----CTTATCCTT	1194
Qy	275	AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla	294
Db	1195	CAGCAAGCCCGGCGACAGAACCTGATGAAGTCAACAGCTGCTAATGGACGGT---ACC	1251
Qy	295	AspIle-----GlyLeuGlyTyrProAsnGlyAsnGln	305
Db	1252	GATTATTGAAAGATCGCGATCCAAATGTTTACGGTCAAGGTTACATCAACCGCAGAACAA	1311
RESULT 11			

148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly 167
1552 CTTACCGAAAGTACGGTGTGTATTTCGTATAGCTGCAGAAATGAAGGTCTCTGCATT 1611
168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
1612 AACATGTTGGAGTCTGTGTGTGGCAACAAAGCAATAACTGTGGAGCTGCTGCA--- 1668
188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
1669 GTGCCATTAAACGTTGGAGTTATGTTCCTCCAGCACTTGGATATCCTGATTACTATGGA 1728
198 -----AsnIleAsnHisValAlaGlnPheSerSerAsgGly 209
1729 TTCATTACTTCCCGCTACACAAGGTT--AGAATAGCATCTCTCAAGCAGAGGG 1785
210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
1786 CCGAGAAATAGATGTTGAATAAAACCAATGTAGTGGCTCCAGGTTACGGAATTTACTCA 1845
230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
1846 TCCCTGCCGATGTGGATTGGCGAGCTGACTTC----- 1878
250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
1879 ---ATGCTCTGGAACCTTCGATGGCTACTCCACATGTCAGCGGTGCGTGCACCTCAT 1935
270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaLeu 289
1936 AGCGGG--GCAAGGCCGAGGGAATATACCAATCCAGATATAATTAAGAAGTTCTT 1992
290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
1993 GAGAGCGGTGCAACCTGCCTTGAGGAGATCCATATACTGGCGAGAAGTACACTGAGCTT 2052
304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
2053 GACCAAGGTCATGCTCTGTGTAACTGATCCCAAGTCTCGGGAATC----- 2097
324 SerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLys 343
2098 -----CTTAGGCTATAAACGGCACC 2118
344 ProLeuLysIleSerLeuValTrpSerAsp-AlaProAlaSerThrThrAlaSerValTh 363
2119 ACTCTCCCAATTGTGTGATCACTGGCGCACAAGTCTCACAGCGACTTGGGAGT---AC 2175
363 rLeuValAsnAspLeu-----AsnLeuValIleThrAlaProAs 376
2176 TTGGTGTGGACGTTATAAGAGTCTCTACGCCAGGAACCTATACCTGACATTGTCGAG 2235
376 nGlyThr--GlnTyrValGlyAsn-----AspPheThrSerProTyrAsnAsp 391
2236 TGGCACATTAAAGTACGTAGGCGCACCGGAGTACAGAATTTTGAGATCTATGCAACTGAG 2295
392 AsnTrp-----AspGly-----ArgAsnAsnValGluAsnVal 402
2296 CCATGGATTAAGCCTTTGTGTCAGTGGAGTGTATTTCTAGAGAACAAATACCGAGTTTGT 2355
403 Phe-----IleAsnAlaProGlnSerGlyThrTyr----- 412
2356 CTTAGGGTGAATATGATGTAGAGGGTCTTGAGCGAGTCTCTATGTTGGAAGGATAATC 2415
413 -----ThrIleGluValGlnAlaTyrAsnValProValGlyPro 425
2416 ATTGATGATCCAAACGCCAGTTATTGGAACGAGATCTTGAACACAATTTGTTATCCC 2475
426 GlnThrPheSer 429
2476 GAGAAGTTCACT 2487

RESULT 12

```

US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

```



```
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 922 AAGGCGGGTACTCCAACTACGGCTCCGCTCTC----- 954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 955 -----GACCTCTTCGGCCCGGTTCTCCATACC 984
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnHisAspSerLysTyr 248
Db 985 TCGGCC-----TGAACCTCAAGCGACTCGCGGACC 1014
QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 1015 AACACCATCTCCGGTACGTGATGGACCCCGACGTGGCGCGCGCGCTC----- 1071
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 1072 -----CACCTCGCCCAACCCCTCGGCCACCCCGTCCAGGTCCGCCACGCGGTGACG 1125
QY 286 LysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 1126 TCGCGCGCCACACCGCGGTCTGACCAACCCCGCGACGGGTCTCGCCCAAC----- 1176
QY 306 GlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaLysValAsnGlu----- 323
Db 1177 -----CGGCTCTGTAGTCTCGCGCGCGGCAC 1203
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 1204 GACCACCTCCGGCGCGCGCTTCGAGAACCCGGTGTACTACACGATCAGCGACAACTCC 1263
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThr 359
Db 1264 ACGTCTGAGTCCCGGTGACGTCTCGCGGTCTCGGCAACCGCGCTCTCGCCCTCGCC 1323
QY 360 AlaserVal-----ThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
Db 1324 GTAGAGTCTCACATCGTCCACACGTATCGGCGACCTCCAGGTCCAGTGTATCGCCGCC 1383
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe-----ThrSerProTyrAsnAspAsnTrpAsp 394
Db 1384 GAGGACGCGGTACAGCTCAAGTCTGACGCGACCGCGCGCGGTCTCGGAC----- 1434
QY 395 GlyArgAsnAsnValGluAsnValPhe-----IleAsnAlaProGln-----SerGly 410
Db 1435 -----AACATCAACACACGTACTCGGTGACGCTCTCTCGGAGCGCGCAACGCGC 1485
QY 411 ThrTyrThrIleGluVal 416
Db 1486 ACGTGGAACTGCGGGTG 1503

RESULT 13
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication NO. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
```

```
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.: 1,17e-21 Length: 59816
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1-COPY (1-434) x US-10-084-846A-1 (1-59816)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 56648 GGGGAGGAGTACGGCGTACGTATCATCGACACCGCGCTC-----CGC 56689
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr-----AlaLeuGlyArgThr 59
Db 56690 ATCACCCACACGACATTCGGCGCGCGGCTCTCTACGGCTACGACGCCATCGACACGAC 56749
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACCGCCAGACGCGCCACGCGCCACGCGACGTCGCGCGCGACGTGCGCGCGCAAC 56809
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 56810 GCC-----TACGGCGTCCGCAAGAGCCCAAGATCGTAGGCTCGCGGTCTGAAC 56860
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 56861 AACTCCGCGCAGGACACCCGCCAGGTGTCTCGCGGTACGTGCGGTGCGCGCGAAC 56920
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 56921 -----GCCGTCAAGCGCGCTCGCCCAACATGTCCCTC 56953
QY 131 GlyAlaValAlaAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArg 150
Db 56954 GCGCGCGCGCCACACGCGCCCTCGACACGCGCTACGCAAC-----GCCATG 57001
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 57002 GCCTCGCGGTACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 57061
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 57062 TCA-----CCCGCACGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 57118
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 57119 AAGCGCGGTACTCCAACTACGGCTCGCTCTC----- 57151
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 57152 -----GACCTCTTCGGCCCGGTTCTCCATACC 57181
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnHisAspSerLysTyr 248
Db 57182 TCGGCC-----TGAACCTCAAGCGACTCGCGGACC 57211
QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 57212 AACACCATCTCCGGTACGTGATGGACCCCGACGTGGCGCGCGCGCGCGCGCTC----- 57268
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 57269 -----CACCTCGCGCCCAACCCCTCGGCCACCCCGTCCAGGTCCGCGCGCGCTGACG 57322
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 57323 TCGCGCGCGCACCGCGGTCTGTACCAACCCCGCGCGCGGTCTCGCGCGCAAC----- 57373
```

```

QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaIleValThrValAsnGlu----- 323
Db 57374 -----CGGCTCTGTACGTCCGGCGGGGCAC 57400
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 57401 GACCACCTCCGGCGCGGCTTCGAGAACACCGGTGACTACACGATCAGCGACAACTCC 57460
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
Db 57461 ACGTTCAGTCCCGGTGAGGCTTCGGGCTCTCCGGCAACGGCGCTCCGCTCCGCC 57520
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuValIleThrAlaPro 375
Db 57521 GTAGAGTCCACATCGTCCACAGTACATCGCGACCTCCAGTCCAGCTGATCGCCCC 57580
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe-----ThrSerProTyrAsnAspAsnTrpAsp 394
Db 57581 GACGGCAGCGGTACACGCTCAAGTCGAGCGACCGGGCGGCGAGTTCGGAC----- 57631
QY 395 GlyArgAsnAsnValGluAsnValPhe-----IleAsnAlaProGln-----SerGly 410
Db 57632 -----AACATCAACACCACTGCTCGGTGACGCTCTCCGGAGCGGCGCAACGGC 57682
QY 411 ThrTyrThrIleGluVal 416
Db 57683 ACGTGGAACTCGGGGTG 57700

RESULT 14
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MOHLENER, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 1,17e-21 Length: 59816
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1-COPY (1-434) x US-10-084-846A-2 (1-59816)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTGACGGCGTACGTATCATGACACCGCGCTC-----CGC 3128
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr-----AlaLeuGlyArgThr 59
Db 3127 ATCAACCCACAGCACTTCGGCGCGCGGCTTCCTACGGCTACGACGCCATCGCAACGAC 3068
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79

```

```

Db 3067 AACACCCCGCAGGACGGCCACCGCCACCGACCGACGTCGCGCGCGGCAAC 3008
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 3007 GCC-----TACGGCGGTCCGCAAGAGGCAAGATCGTAGCGGTCCGCGTCTGAAC 2957
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 2956 AACTCCGGCCAGGACACACCGCCAGGTGTCGCCGCGCATCGACTGGTGGTCCGCCGAAC 2897
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 2896 -----GCCGTCAAGCGCGCGTCCGCAACATCGCTCCCTC 2854
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspTyrValArg 150
Db 2863 GCGCGCGCGCGACACGCGCCCTCGACACGGCCGTACGCAAC-----GCCATG 2816
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 2815 GCCTCCGGCGTCACTTCGCCGTCGCGCGCAACGAGTCGACCAACGCTCCACGAGG 2756
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 2755 TCA-----CCCGCACGCGTCACCGAGGCCATCACCGTCCGCGCGACGACGCTCGGCGCC 2699
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 2698 AAGCGCGCTACTCAACTACGCTCGCTCCTC----- 2666
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 2665 -----GACCTCTTCGCCCGCGGTCTGCTCCATCACC 2636
QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyr 248
Db 2635 TCGGCC-----TGGAACCAAGCACTCGCGGACC 2606
QY 249 AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 2605 AACACCATCTCCGGTACGTGATCGACACCGCGACGTCGGCGGCGCGCGCTC----- 2549
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 2548 -----CACCTCGCGCGCAACCTTCGGCCACCCCTCCCGAGTCGCCACGCGCTGACG 2495
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 2494 TCCGCGCGCACACCGCGGTCTGTCACCAACCCCGCACGCGCTCGCCCAAC----- 2444
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
Db 2443 -----CGGCTCTGTACGTCCGCGCGGCGGAC 2417
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 2416 GACACCTCTCCGGCGCGCGGTTCGAGAACACCGGTGACTACACGATCAGCGACAACTCC 2357
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
Db 2356 ACGGTTCGATCCCGGTGACGCTCTCCGGCGCTCTCCGGCAACCGCGCTCCGCGCTCCGCC 2297
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
Db 2296 GTAGAGTCCACATCGTCCACAGTACATCGCGGACCTCCAGGTCCAGCTGATCGCCCC 2237
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe-----ThrSerProTyrAsnAspAsnTrpAsp 394
Db 2236 GACGGCAGCGGTACACGCTCAAGTCGATACGGCACCGCGCGCGAGTTCGGAC----- 2186
QY 395 GlyArgAsnAsnValGluAsnValPhe-----IleAsnAlaProGln-----SerGly 410

```

```

Db 2185 -----AACATCAACACACGACTCTCGTGAACGCTCTCTCGGAGCGGCCAACAGGC 2135
Qy 411 ThrTyrThrIleGluVal 416
Db 2134 ACGTGGAAACTGCGGGTG 2117

RESULT 15
US-09-927-827-33
; Sequence 33, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 33
; LENGTH: 3788
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2788)
US-09-927-827-33

Alignment Scores:
Pred. No.: 1,01e-20 Length: 3788
Score: 282.00 Matches: 112
Percent Similarity: 40.86% Conservative: 49
Best Local Similarity: 28.43% Mismatches: 113
Query Match: 12.55% Indels: 120
DB: 10 Gaps: 19

US-09-985-689a-1-COPY (1-434) x US-09-927-827-33 (1-3788)
Qy 3 ValAlaArgGlyLeuValAlaAspValAlaGlnSerTyrGlyLeuTyrGlyGln 22
Db 1739 GTCTCGGGCAACGTGTGTGATGTGGATGCAGCAGCAGCGCGTATCGCCAGTGGGCG 1798
Qy 23 GlyGlnIleValAlaValAlaAspThrGlyArgAsnAspSerMet 42
Db 1799 CAGATCGCTGGGGTCTCTGGATACCGGATT-----GGGGCCGCG 1840
Qy 43 HisGluAlaPhe-----ArgGlyIleThrAlaLeuTyr----- 54
Db 1841 CATCCGCAATTTCTTCGTCCGCGCGAGCGACATGTGGTGGCGCAATGGGATTCACG 1900
Qy 55 -----AlaLeuGlyArgThrAsnAsnAlaAsnAsp 64
Db 1901 CGGGCGGGCGGCCAAGCGGTGACCGCGCGCGATGGCAACGGTTCACCGCACTCGAC 1960
Qy 65 ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
Db 1961 GGGCAGCGCCATGGCAGCAGCATCGCGCGCATCATCGCGGTGTTCGCCGGCGCGGTG 2020
Qy 79 -----AsnGlySerThrAsnLys-----GlyMetAlaProGlnAla 90
Db 2021 CCTGATCCAGTGGCAAGCGGGGAACTGCTGGAATTCGCGCATGGCGCGGAGAGC 2080
Qy 91 AsnLeuVal---PheGlnSerIleMetAspSerGlyGlyLeuGly----- 106
Db 2081 CAATCTATGGCTTCAAGGTGTGGACGACGCGCGCAACGCGCGCGATTCTGTGGATGATC 2140
Qy 107 ---LeuProSerAsnLeuGlnThrLeuPheSerGlnAla----- 118
Db 2141 AAGGCAGTGCAGCAGTGGCGGCTCTCAACAGCGGTGCGCGAGCTGTGTGATCCAGCGC 2200
Qy 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaValAsnGly 136

```

```

Db 2201 GTCAATCTCAGCTGGCGGGTACTTCGATCCGGAAGCTACGGCTGTGGCTTC----- 2254
Qy 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
Db 2255 -----ACGCGGTATGCAACGAGTTGCGCGGTATTGCGCGCAGCGCGTACTGTA 2305
Qy 157 LeuPheAlaAlaGlyAsnGluGly-----ProAsnGlyGlyThr--- 169
Db 2306 GTGTGGCGCGCGCAACGAGGCTTGGCTGGCTGATGCAGAACGACGCGCGCACCTAT 2365
Qy 170 -----IleSerAlaProGlyThrAlaLysAsnAlaIleThrVal 182
Db 2366 CCGCGCAACATGATCTGTCGATCAGCGATCCGGCAATCTGGAGACGCGATCGTGGTG 2425
Qy 183 GlyAlaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 2426 GGATCGGTGCACAGAGCAGCCCGCAACAAATACGGC----- 2461
Qy 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProaspValMet 221
Db 2462 GTGTGTAATTTTTCATCCCGCGCGCGCGCGATGGCGCTCCAAACCTGATGTGTGTC 2521
Qy 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProaspSerSerPheTrp 241
Db 2522 GCGCGCGCGCAAGAAAGATCTCTGCTTACTACGGCTTCGACCCG----- 2566
Qy 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
Db 2567 CGGACCCGTCACGCTGATGTCGAGATGAGCGGCGCACCGATGCGCGCATGTG 2626
Qy 262 AlaGly-----Asn 264
Db 2627 TCGGCGTGTGCGCGGTTTTTATCCGACGCGCGAGTTTCATCGGCTTCGCGACCGG 2686
Qy 265 ValAlaGlnLeu-----ArgGluHisPheValLysAsn 275
Db 2687 GTCAAGCACTGCTGTGGACACTGCACCGACCTGCAGCGCGATGTTACGTGCGAGGC 2746
Qy 276 ArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAsp 295
Db 2747 AGGGCGGTG-----CGAAATTTGATCGCG-----ATGCTTGGAGAGACGTGA 2788
Qy 296 IleGly-----LeuGlyTyrProAsnGlyAsnGlnGly 306
Db 2789 TTCGGAGTCCGGATTTGGGATTCGCAACCGCGTGTGGC 2830

```

Search completed: March 16, 2004, 01:21:14
Job time : 7202 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 20:30:22 ; Search time 2385 Seconds
(without alignments)
5434.040 Million cell updates/sec

Title: US-09-985-689A-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPQTFLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DBV=xlh
-Q=/cgm2_1/USPTO.spool/SWOPE985689/runat_10032004_112806_19374/app_query.fasta_1.583
-DB=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=DC -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=SWOPE985689@cgm2_1_2135@runat_10032004_112806_19374 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	246	10.9	1605	13	BQ622771	BQ622771 CC Contig
2	241	10.7	640	12	BJ395336	BJ395336 BJ395336
3	226.5	10.1	4198	11	AK029048	AK029048 Mus muscu
4	224.5	10.0	594	12	BJ393752	BJ393752 BJ393752
5	218.5	9.7	532	29	TAJ19G10P	TAJ19G10P T. brucei
6	207.5	9.2	2141	13	BQ142519	BQ142519 Contig M
7	200	8.9	508	28	AQ652212	AQ652212 Sheared D
8	199.5	8.9	601	12	BJ387574	BJ387574 BJ387574
9	195.5	8.7	1002	29	CNS0606B	AL393417 T3 end of
10	195	8.7	771	14	CA320325	CA320325 UI-M-FW0-
11	191	8.5	633	12	BJ369190	BJ369190 BJ369190
12	190	8.5	718	12	BI750157	BI750157 Fg02_1090
13	186.5	8.3	4662	11	BC060627	BC060627 Mus muscu
14	185.5	8.3	574	29	TAJ15H10P	AL490202 T. brucei
15	182.5	8.1	614	9	AJ273402	AJ273402 AJ273402
16	181	8.1	716	28	BZ893395	BZ893395 HL2_0177
17	180	8.0	641	12	BJ393925	BJ393925 BJ393925
18	180	8.0	665	13	BQ770462	BQ770462 UI-M-P10-
19	179.5	8.0	3091	11	BC011275	BC011275 Mus muscu
20	177.5	7.9	650	9	AJ274038	AJ274038 AJ274038
21	177.5	7.9	681	14	CB690041	CB690041 CBST-54-B
22	177	7.9	675	14	CF727824	CF727824 UI-M-HB0-
23	176	7.8	530	29	CNS010PO	AL153820 Anopheles
24	176	7.8	576	14	CD295943	CD295943 StrPu691.
25	176	7.7	580	28	BZ424995	BZ424995 100023066
26	172.5	7.7	583	9	AJ273947	AJ273947 AJ273947
27	172.5	7.7	593	9	AJ273918	AJ273918 AJ273918
28	172.5	7.7	601	9	AJ273921	AJ273921 AJ273921
29	171.5	7.6	573	14	CA937626	CA937626 sav42b10.
30	171.5	7.6	601	9	AJ273050	AJ273050 AJ273050
31	170.5	7.6	712	14	CD311344	CD311344 StrPu691.
32	170.5	7.6	1029	29	CNS071DW	AL424794 T7 end of
33	170.5	7.6	895	13	BQ216158	BQ216158 AGENCOURT
34	170	7.6	545	13	BU575479	BU575479 TGESTzydB8
35	169.5	7.5	545	13	BU575479	AJ273185 AJ273185
36	169.5	7.5	604	9	AJ273185	CF737198 UI-M-HD0-
37	169	7.5	794	14	CF737198	AY107161 Zea mays
38	169	7.5	1572	11	AY107161	BJ365857 BJ365857
39	168.5	7.5	609	9	AJ272712	AJ272712 AJ272712
40	168.5	7.5	937	12	BG246418	BG246418 602360428
41	168.5	7.5	831	14	CA510555	CA510555 UI-R-FU0-
42	168	7.5	1050	29	CNS0780L	AL433387 T3 end of
43	168	7.5	564	28	AQ651427	AQ651427 Sheared D
44	167.5	7.5	564	28	AQ651427	AJ274218 AJ274218
45	167.5	7.5	610	9	AJ274218	

ALIGNMENTS

RESULT 1
BQ622771
LOCUS CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
DEFINITION CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
ACCESION BQ622771.1 GI:21649940
VERSION BQ622771
KEYWORDS EST.
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

Qy 49 LysleThrAlaLeuTyraLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
 Db 142 GTTGTACT-----TATATTACTACATGACATGAAGATTATGTAATGTCAT 192
 Qy 69 GlyThrHisValaGlySerValLeuGlyAsnGly-----Ser 81
 Db 193 GGTACACATGTTGTGCTCTGCAGCAGGTACTCCAGAGATTCTTCATGGCGCTATTTC 252
 Qy 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
 Db 253 TCATTAGTGTCTGCACTGACGCAAGATTGCATTATGAT---CCTTCATCTGA 309
 Qy 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyraAla 121
 Db 310 AGTTCCTGAACCAACACCCAGAGATTACAGTCAAAATGATCAAAACCATATATGATGCA 369
 Qy 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGlyAla 137
 Db 370 GGTCAAGAGTACATGTTGTTGGGTCTGTATCTTTGCAAGGTTATTTATGTTGTT 429
 Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIle 156
 Db 430 TATTCGATGATGCTGGTGTATTGATGATCTCTCTATGATGATCCAGCAATCTCTATA 489
 Qy 157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
 Db 490 CTAAGAGCTGCTGTAAT---AAGAGCTATTGTCATCTTTATGATCAAGCAACAGCT 546
 Qy 177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer----- 194
 Db 547 AAAATGCAATTACAGTTGGTGTGTCGACCAACAGCTCATGTAAATATGTTGTCAGATGCA 606
 Qy 195 -----TyrAlaAspAsnIleAsn 200
 Db 607 TTGGAATATTATGATTCTTCAGATAATGCTAAT 639
 RESULT 3
 AK029048
 LOCUS
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473248M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.
 ACCESSION AK029048
 VERSION AK029048.1 GI:26325017
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 2
 20499374
 10349636
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL MEDLINE
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayaehida, K., Hayateu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
 COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 URL: http://location/Qualifiers
 1. 4198
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:473248M11"
 /db_xref="MGI:2391080"
 /db_xref="taxon:10090"
 /clone="473248M11"
 /tissue_type="skin"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"
 417. 3575
 /note="unnamed protein product; membrane-bound transcription factor protease, site 1 (MGD|MGI:1927235, GB|NM_019709, evidence: BLASTN, 99%, match=3782) putative"
 /codon_start=1
 /protein_id="BAC26263.1"
 /db_xref="GI:26325018"
 /translation="MKGVSTWLLVLVLVLLCGKRLHGLDGLTRALEKAPCPSCSHLTLEKVEFSTVVEYIYVAFNGYFTAKARNSFISALKSESEVENWRIIPRNPSSDIYPSDFEVIQIKKQKAGLLTLEDHFNKVTQPKRFRSLKFAESNPVFNPNFWSQKQWSSRPLKRASLSIGSGFHWATGRHSRLRLRAIPQVATQADVLQWGYTGANRVAVFDTLSKTHFKNKVKTNTWRTUDDGLGHGTGVAGVIAVMRECOGAPDAELHIFVETNQVSYTWFLEDAFNVALIKKMDVNLISIGGDFDNDHDFVDKWTANNVIMVSALNDGPIYGLNPNADMDVIGGIDFENIARPPSRGTTWELPGYGRVKEDIVTYGAVRSGVGGCGRSLSVASVAVAGATLLVSTVQKRELNPASVQKALIASARELPVNMFEQHGKLDLRLAYQLSSYQPSLSYDITDECPMPFYCSQPIYGRGPTIVNTILNGMGTGRIVDKPEWRPYLPQNGDNIETVAFSYSVLWFWGTYLAISI

CDS

SVTKKAASWEGIAQGHIMITVASPAETELHSGAERTSTVKLPKIKVITPTPRSKRVL
 WQYNLRYPGFFPRDLNRMKNDPLDWNQDHVHTNFRDMYQHLRSMGYFVEVLGAPF
 TCEDATQVCTGLLVDSBEYFPEETAKLRDNDVNLGLSVIFSDWNTYVMSKVKFYDE
 NTRQWMDPTGGANIPALNELLSVNMWFSGLXEGEFVLNHDHMYASGCSIAKFFE
 DQVITQIFKQDGLSVLQKQETAVENVILGLYOIPSGGSEIVLYGDSNCLDSDHQ
 KQCFMLLDALLOTYSVITPPSLSHSGNRQRPSPGAGLAPPERMEGNHLHRYSKVLEA
 HGDGPKRPLPACPLNLTAPSNLWKHKQLLSLDLDKVLNLPFNRSRNPQV
 RPLSPESGAMDIPOGIMPRYNOBGVQTIPIVFAELGAWVALAFVWQISKAKSRPKR
 RRPRAKRPQLAQCAHPARTPSV"
 4180..4185
 /note="putative"
 polyA_signal
 4198
 /note="putative"
 polyA_site

ORIGIN

Alignment Scores:
 Pred. No.: 7,31e-11 Length: 4198
 Score: 226.50 Matches: 116
 Percent Similarity: 41.32% Conservative: 65
 Best Local Similarity: 26.48% Mismatches: 160
 Query Match: 10.08% Indels: 97
 DB: 11 Gaps: 25

US-09-985-689A-1-COPY (1-434) x AK029048 (1-4198)

Qy 8 ValiysalaaspValalaginSerserTygLyLeuTyrglycnglyGlnileValala 27
 Db 1005 CTCAGGACATGCTGGT---CAGATGGGATACAGGTGCTTAATGTCAGATTGCT 1061
 Qy 28 ValalaspThrglyLeuaspThrglyArgAsnaspSerserMethisGluAlaPheArg 47
 Db 1062 GTTTTGTACTGGCTC-----AGTGAGAGCATCCGCAATTTTAA 1103
 Qy 48 GlylylsileThralaleuTyralaleuGlyArgThrasnAsnAlaasn----- 63
 Db 1104 AAT-----GTGAAGGAGAGAACCAATCGACCAATGAGCGGACCCCTG 1145
 Qy 64 ---AspThrasnGlyHisGlyThrisValalaglyserValleuGlyAsnGlyserThr 82
 Db 1146 GATGATGGCTAGGCCATGACATTCGTTGCAGGT---GTGATTGCCAGCATGAGGAG 1202
 Qy 83 AsnlyGlyMetAlaProGlnAlaAsnLeu---ValpGlnSerserileMetaspSergly 101
 Db 1203 TGCCAAAGATTGCTCCAGATGACAGCTGCACATCTTCAGGGTCTTTTACCACAAAT--- 1259
 Qy 102 GlyglyLeuGlyGlyLeuProSerAsnLeuGlnThrlPheSerGlnAlaTySerAla 121
 Db 1260 -----CAGGTCTCTACACATCTGTTCTGATGCGTCCTCAACTAT 1301
 Qy 122 GlyAlaArgTleHisThrasnSerTrpGlyAlaAlaValasnGlyAlaTyThrlThrApp 141
 Db 1302 GCCATCCTAAGAAGATGGAGCTTCTCAACCTTAGCATCGTGGCGCCGACTTCATGGAT 1361
 Qy 142 SerArgAsnValaspAspTy-----ValArgLyAsnaspMetThrlleLeuPheAla 159
 Db 1362 CATCCGTTTGTGCAAGGTGGGAATTAACAGCTAACATGTAATTAATGTTTCTGT 1421
 Qy 160 AlaGlyAsnGluGlyProAsnGlyGlyThrlSeraAlaProGlyThralalysAsnAla 179
 Db 1422 ATTGGCAATGATGACCTCTCTATGGCAGCTCTGAATAACCCCTGTCATCAGATGGATGT 1481
 Qy 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlyserTyralaAspAsnIle 199
 Db 1482 ATTGGATGGGTGGCATTGAC-----TTTGAGATACATC 1517
 Qy 200 AsnHisValAlaGlnPheSerserArgGly-----ProThrllysAsp 213
 Db 1518 -----GCTCGCTTTCTTCAGGGGAATGACTACTCTGGGAATTAACAGGAGGCTAT 1568
 Qy 214 GlyArgTleIysProaspValMetAlaProGlyThrlPheIleLeuSeraAlaArgSerser 233
 Db 1569 GGTGCTGTGAAGCTGACATTGTC-----ACCTATGGTGTGAGTGGCGGGTTC 1619

Qy 234 LeuAlaProaspSerserPheTrpAlaasnHisaspSerserLyTyAlaTyMetGlyGly 253
 Db 1620 GGTGTGAAAGGGGGCTGC-----CGTGACACTTCTCAGG 1652
 Qy 254 ThrSerMetAlaThrProileValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 Db 1553 ACAGTGTGCTTCCCGAGTGGTCTGGGCGCGTCACTTGTGTAGTAAAGCACAGTACAG 1712
 Qy 274 LyAsnArgGlyIleThrProIysProSerLeuLeuLyAlaAlaLeuIleAlaGlyAla 293
 Db 1713 AAGCGGAGCTGGT-----AATCTGCCAGTGTGAAGCAAGCTTGTAGCGTCAGCC 1766
 Qy 294 AlaaspIleGlyTyProAsnGlyAsn-----GlnGlyTrpGlyArgVal 310
 Db 1767 CGGAGACTT-----CTGGGGTCAACATGTCGAGCAAGTTCATGCAAGTTG 1814
 Qy 311 ThrleuAspLysserLeuAsnVal---AlaTyValasnGluSerserSerserLeuSerThr 329
 Db 1815 GATCTGCTGGAGCTTATCAGATCTCAGCAGCTATAAACCGCAGGCAAGCTTGTGCTCT 1874
 Qy 330 Ser-----GlnLyAlaThrTySerPheThrAlaThralaGlyLysProLeu 345
 Db 1875 AGCTACATCAGCTGACTGCTGCTCCTACATGTTG---CCCTACTGCTCCAGCCT--- 1928
 Qy 346 LysIleSerLeuValTrpSerAspAlaProAlaSerThrlThrAlaSerValThrLeuVal 365
 Db 1929 -----ATCTACTATGAGGAATGCCA-----ACAATCGTTAATGTCAACATCTCT 1973
 Qy 366 AsnaspLeuAsnLeu-----ValIleThrAlaProAsnGlyThrlGlnTyValGly 382
 Db 1974 AATGGATGGCGTGCAGGAAAGATTGTGGTAAGCTGAGTGGCGGACCTATTATTA--- 2030
 Qy 383 AsnaspPheThrSerProTyAsnaspAsnTrpAspGlyArgAsnAsnValGluAsnVal 402
 Db 2031 -----CCACAGAAATGGA-----GACACATTAAGTGAAGTGGCC 2060
 Qy 403 Phe-----IleAsnAlaProGlnSerserGlyThrlThrIleGluVal 416
 Db 2061 TTCTCTACTCTCAGTGTGTGGCCCTGTCAGTTTACCTTGCCATCTCCATT 2114
 RESULT 4
 BU393752 594 bp mRNA linear EST 08-MAR-2002
 LOCUS BU393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
 DEFINITION discoideum cDNA clone dds32b16 5', mRNA sequence.
 BU393752
 ACCESSION BU393752.1 GI:19304838
 VERSION BU393752
 KEYWORDS EST
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 594)
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..594
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds32b16"
 /sex="mat A"
 /dev_stage="Slug stage"
 /clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Alignment Scores:
 Pred. No.: 4.3e-12 Length: 594
 Score: 224.50 Matches: 59
 Percent Similarity: 50.80% Conservatives: 36
 Best Local Similarity: 31.55% Mismatches: 59
 Query Match: 9.99% Indels: 33
 DB: 12 Gaps: 9

US-09-985-689A-1-COPY (1-434) x B03993752 (1-594)

QY 49 LysilleThralaleuTyralaleuGlyArgThrAsnAlaAsnAspThrAsnGlyHis 68
 Db 21 AAAGTTGTAACATATATACCATCAACAGCAGATAGTGAATAAGTGGTGCAC 80
 QY 69 GlyThHisValalaglySerValLeuGly 81
 Db 81 GGTACACATATTTGTTGTCGAGCAGGTACTCCAGAGGATTTCTTCATGTTAATATTTCA 140
 QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerileMetAspSerGly 101
 Db 141 TCATTAGTGTCTTGCACACTGATGCAAGATTGCATTC-----TTTGATTGGCA 191
 QY 102 GlyGlyLeuGlyGlyLeu-----ProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
 Db 192 AGTGGTTTCATCAAGTTTGACACCTCCATCGGATTTGAAACAATTTATCAACCATTTATAT 251
 QY 120 SerAlaGlyAlaArgileHisThrAsnSerTyrGlyAla-----AlaValAsn 135
 Db 252 GAGCAGGTCAAGAGTGCATTTGTAATCTTGGGGTCTGTATCAGTAGAGGGGTATACA 311
 QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspTyrValArgLysAsn---AspMet 154
 Db 312 GGTAGTTATTATCATCAGACACTGTTCAATGATGATTCCTTTTCACTCATCCAGATTTTC 371
 QY 155 ThrileuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrileSer-----Ala 172
 Db 372 ATCATTTTATAGCTGCTGTGTAAC-----AACGACATATCATCTACTACTACT 422
 QY 173 ProGlyThrAlaLysAsnAlaileThrValGlyAlaThrGluAsnLeuArg----- 189
 Db 423 CAATCCACTCAAGAATGTTATTACCGTTGGTGTCTCATCAACAATTCATGAAATAT 482
 QY 190 -----ProSerPheGlySerTyrAlaAspAsnileAsn-----His 201
 Db 483 TTAAGTATGATGTCATTAATATATAATATCAATCATCTGTCATATAATAATCAAGAGTTA 542
 QY 202 ValAlaGlnPheSerSerArg 208
 Db 543 ATATGCTATTCGATACGAGA 563

RESULT 5
 TA319G10P 532 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
 DEFINITION genomic survey sequence.
 ACCSSION AL492464
 VERSION AL492464.1 GI:11867408
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 532)
 Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 nh@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TBEU927/4 GMTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
 Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

Location/Qualifiers
 1..532
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TBEU927"
 /db_xref="taxon:5691"
 /clone="319g10"

ORIGIN

Alignment Scores:
 Pred. No.: 1.44e-11 Length: 532
 Score: 218.50 Matches: 61
 Percent Similarity: 48.37% Conservatives: 28
 Best Local Similarity: 33.15% Mismatches: 68
 Query Match: 9.73% Indels: 27
 DB: 29 Gaps: 6

US-09-985-689A-1-COPY (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerileMetAspSerGlyGlyLeuGly----- 105
 Db 2 GTAATGCCAAGTATAAAGTGTCTCCAGAGGGGCGAAGATTCTTCAGGGGTGGGGTGGC 61
 QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
 Db 62 CATCCAGTCAGAGCTTGTCTCTCCCCACGACGTTACTCAAAATTATTCGTCGCGGTATAT 121
 QY 120 SerAlaGlyAlaArgileHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 139
 Db 122 GCGCTGGAGCCCGTGTGTTCTCAAACTCGTGGGGTTTGTGTCTCCCTCGAGTATCT 181
 QY 140 ThrAspSerArgAsnValAspTyrValArgLys---AsnAspMetThrileLeuPhe 158
 Db 182 GCTGTGGAAAGGATATGATGATGTTGCGAGTAGTATTGACGATGCGTACTTATCTTC 241
 QY 159 AlaAlaGlyAsnGlyProAsnGlyGlyThrileSerAlaProGlyThrAlaLysAsn 178
 Db 242 TCCACTGGCAACAGTATCCAAAGATGGC-----CTAATGACTCGGTGCGGTGAAGAAC 295
 QY 179 AlaileThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
 Db 296 GTGATGTCGTGGGTGCACACAAAACGTG-----TTTGACGCTTCGAAAGAC--- 343
 QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 344 -----ATTGTTCTTCTCGTATCGTATCGTATCGTATCGTATCGTATCGTATCGT 397
 QY 219 AspValMetAlaProGlyThrPheileLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 398 GATCTGTGCGTCCCGGGAAGAGGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 457
 QY 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 458 AAA-----CAATGTAAGTGTGGCAAGCGGNGTTCATCGATGCGCAACT 502

259 ProfileValala 262
 503 GCGGCCGTCGCG 514

RESULT 6

BQ142519/c

LOCUS

BQ142519

2141 bp

mRNA

linear

EST 24-APR-2002


```

http://image.llnl.gov
(This clone was contributed by the Brain Molecular Anatomy Project
(BMAP))
Seq primer: pyx-5.
Location/Qualifiers
1.771
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stages="embryo 13.5,14.5,16.5,17.5dpcc"
/lab_host="DHI0B (T1 phage resistant)"
/clone_lib="NTH BMAP FMO"
/note="Organ: Brain; Vector: pyX- Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is ACGGACAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
FEATURES
source

```

ORIGIN

Alignment Scores:		
Pred. No.:	6,26a-09	Length:
Score:	195.00	Matches:
Percent Similarity:	41.8%	Conservative:
Best Local Similarity:	28.5%	Mismatches:
Query Match:	6.68%	Indels:
DB:	14	Gaps:
		771
		82
		38
		109
		58
		13
US-09-985-689A-1-COPY (1-434) x CA320325 (1-771)		

Qy	8	VallysAlaAspValaAlcInSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla	27
Db	40	CTCGAGGCAGATGTCGTGG---CAGATGGCATACACAGGTGCTAATGTCAGAGTTGCT	96
Qy	28	valAlaAspThrGlyLeuAspThrClyArgAsnAspSerSerMethHisGluAlaPheArg	47
Db	97	GTTTTGTGATCTGGGCTC-----AGTGAGAAGCATTCGCATTTTAAAG	138
Qy	48	GlyLyIsIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn	63
Db	139	AAT-----CTGAAGGAGAGAACCACTGGACCAATGAGCGGACCCCTG	180
Qy	64	---AspThrAsnGlyHisClyThrHisValAlaGlySerValLeuGlyAsnGlySerThr	82
Db	181	GATGATGGGCTAGGCCATGGCCATTCGTTGCAGGT---GTGATGCCAGCATGAGGGAG	237
Qy	83	AsnLySglMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly	101
Db	238	TGCCAGGATTGCTCCAGATGCAGAGTCGCACATCTTCAGGGTCCTTTACCAACAT---	294
Qy	102	GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrIleuPheSerGlnAlaTyrSerAla	121
Db	295	-----CAGGTGCTTACACATCTTGGTTTCTGGGATGCGCTTCAACTACT	336
Qy	122	GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThrAsp	141
Db	337	GCCATCTTAAGAAGATGAGCGTTCTCAACCTTAGCATCGTGGCGCGCATTCATCGAT	396
Qy	142	SerArgAsnValAspAspTyr-----ValArgLyAsnAspMetThrIleLeuPheAla	159
Db	397	CATCCGTTTGTTCACAAGGTGGGAATTAACAGGTAACTAATGTAATATGTTGTTCTGCT	456

Qy	160	AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla	170
Db	457	ATTGCCAATGATGACCTCTCTATGGCACCTCTGAATAACCTCTCTGATCAGATGGATGTG	516
Qy	180	IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle	199
Db	517	ATTGAGATGGGTGGCATTGAC-----TTTGAAGATAACATC	552
Qy	200	AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp	213
Db	553	-----GCTCGCTTTCTTCAGGGGAATGACTACCTGGGAATTACGAGGAGGCTAT	603
Qy	214	GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer	233
Db	604	GGTCGTGTAAGCCCTGACATTGTC-----ACCTATGGTGCTGGAGTGGCGGGGTCC	654
Qy	234	LeuAlaProAspSerSerPheTirAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly	253
Db	655	GGTGTGAAGGGGGGTGC-----CGTGCACCTCTCAGGG	687
Qy	254	ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal	273
Db	688	ACCAGTGTGCTTCCCACTGGTCTGCTGGGGGGCTCACCTTGTTAGTANGACACAGTACAG	747
Qy	274	LysAsnArgGlyIleThrPro	280
Db	748	ARGGGGAGCTGGTGAATCCT	768

RESULT 11

Accession	Gene	Length	Library	EST
BJ3369190	Dictyostelium discoideum	633 bp	linear	EST 08-MAR-2002
LOCUS				
BJ3369190	Dictyostelium discoideum	633 bp	linear	EST 08-MAR-2002
DEFINITION	Dictyostelium CDNA clone ddc49116 5', mRNA sequence.			

ACCESSION BU369190
 VERSION BU369190.1 GI:19278573
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum

ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
1 (bases 1 to 633)
REFERENCE Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
AUTHORS Full length cDNA of Dictyostelium discoideum at the
TITLE stage

Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp

```

FEATURES
source
Location/Qualifiers
1. 633
/organism="Dictyostelium discoideum"
/mol type="rRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc4916"
/sex="mat A"
/dev_stage="Culmination stage"
clone_1ib="Dictyostelium discoideum"

```

OPTICIN

Alignment Scores:	1.14e-08	Length:	633
Pred. No.:	Score:	Matches:	57
	191.00	Conservative:	29
	Percent Similarity:	Mismatches:	49
	41.55%	Indels:	72
	Best Local Similarity:	Gaps:	6
	27.54%		
	Query Match:		
	8.50%		
	Da:		
	12		

US-09-985-689A-1-COPY (1-434) X BJJ369190 (1-633)

QY 136 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMet 154
 Db 6 GGTGGTATTCCGATGATGCTGGTGGTATTGATGCAATCCCTATGATGATCCAGAAATC 65
 QY 155 ThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGly 174
 Db 66 TCTATCTAAAGAGCTGCTGGTAAT---AACGAGCTATTGGCATCTTTATTAGCTCAAGCA 122
 QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
 Db 123 ACAGCTAAATGCAATACATGCTGGTCTGAGCAACAGCTCATGTAATATTATGTGCA 182
 QY 195 -----TyrAlaAspAsnIle----- 199
 Db 183 GATGCAATGGAATATTGATTCTCAGATATGCTAATTTTCAAAGACCATGTTTATTC 242
 QY 199 ----- 199
 Db 243 GATAAGAGCTATTGTAATTATACAGCCGCTAAATGCTGCTAGAGGTTTCAAATGTTAAA 302
 QY 199 ----- 199
 Db 303 GGTTCACAAATTATGTTGTCAGCATCTATTAAACAAATGCGATCGGATTCATTACAACA 362
 QY 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLys 212
 Db 363 CAACCTCAATTTTATATGAAATATATGGGATCATCTCATCAAGGGTCCACACAT 422
 QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
 Db 423 GATGTAAGATTGAACCTGATATAGTTCACCTGTTGTAATATATACATCGGAGATCA 482
 QY 233 Ser-----LeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 246
 Db 483 AATGTTGAGAAATTCACAGACCAATGTGGTGTCTTTTA-----CCAAATCCAAAT 536
 QY 247 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
 Db 537 GGTCTAATGTCATATCTGGTATCATGATGGAACACCATTCGGCAACAGCAGCAACA 596
 QY 267 GlnLeuArgGluHisPheVal 273
 Db 597 ATTCTTAGACAAATATTAGTT 617

RESULT 12
 BI750157 718 bp mRNA linear EST 25-SEP-2001
 LOCUS Fg02_10508_R Fg02_AAFc_ECORC_Fusarium_graminearum_mycelium
 DEFINITION Gibberella zeae cDNA clone Fg02_10508, mRNA sequence.
 ACCESSION BI750157
 VERSION BI750157.1 GI:15771959
 KEYWORDS EST.
 SOURCE Gibberella zeae

ORGANISM Gibberella zeae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE 1 (bases 1 to 718)
 HARRIS, L.J., GLASSCO, T., ROCHELLEAU, H., ALLARD, S., CHAPADOS, J.,
 COURCOUX, P., DE MOORS, A., HATTORI, J.I., OUELLET, T., ROBERT, L.S.,
 SINGH, J.A., SPROTT, D. and TINKER, N.A.
 Expressed Sequence Tags from Fusarium graminearum mycelium
 Unpublished (2001)

TITLE Gibberella zeae
 JOURNAL Contact: Harris, Linda J.
 COMMENT Eastern Canada and Oilseed Research Centre
 Agriculture and Agri-food Canada
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
 CANADA

Tel: (613) 759-1314
 Fax: (613) 759-6566
 Email: harris.lj@em.agr.ca.
 Location/Qualifiers
 1. 718

FEATURES
 source

ORIGIN

Alignment Scores:
 Pred. No.: 1.77e-08 Length: 718
 Score: 190.00 Matches: 66
 Percent Similarity: 44.05% Conservatives: 34
 Best Local Similarity: 29.07% Mismatches: 81
 Query Match: 8.46% Indels: 46
 DB: 12 Gaps: 10
 US-09-985-689A-1-COPY (1-434) x BI750157 (1-718)
 QY 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThr 65
 Db 9 TTCGAGGTCGTCTGCTCAGGCTGTCTACACTGCTTCAGCGCGGCAGAACGCT---GACACC 65
 QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLysGly 85
 Db 66 AAGCGTCACGGAACACTCAGTTGCTGCGACTATTGCCGGAAG-----ACATACGCT 116
 QY 86 MetAlaProGlnAlaAsnLeu-----ValPheGlnSerIleMetAspSerGly 101
 Db 117 GTTGCCAAAGAAGGCCACCATCCAAAGCTGTCAAGGTCTTCAG----- 158
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer--- 120
 Db 159 ---GGTAGTTTCATCCAGACCTCCATCATCTCGCTGGCTTCACTGGGCTGCAACAGCAG 215
 QY 121 -----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
 Db 216 ATCATCTCCAAAGGCCGACCAACAGACVTCAGTCGTCATATGTCCTCGCGCGTGGTTAC 275
 QY 139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
 Db 276 TCTGCTTCCTTCAACACACGCTGTCGAGTCTKCTCCAGCTCCGGTATTATCTCTGCCAAT 335
 QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
 Db 336 GCTGCCGGTAACGATGTCGCCAAGCGTCCCAACACTTCT---CTGCCTCTTCTCCAGC 392
 QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
 Db 393 GCCAACAATCTGCGTGGTCCATTGACAGC-----AACTGGGCC----- 428
 QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 429 -----ATTCKCTCGTACTCAACTACGTTACCGTTCTC----- 461
 QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 462 GATATCTTTGCTGCTGACAGCAGCGTCTCTCCGCC----- 497
 QY 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 498 -----TGGTACACCAACAGCAGTCCACCAACACCATCAGCGKACSTCNTGGTACT 551

/organism="Gibberella zeae"
 /mol_type="mRNA"
 /strain="DAOM 180378"
 /db_xref="taxon:5518"
 /clone="Fg02_10508"
 /tissue_type="Mycelial tissue"
 /dev_stage="Asexual"
 /lab_host="E. coli (Sure cells)"
 /clone_lib="Fg02_AAFc_ECORC_Fusarium_graminearum_mycelium"
 /notes="Vector: Bluescript SK+/XhoI-ECORI; Site 1: EcoRI;
 Site 2: XhoI; Mycelial tissue was collected from v8 agar
 plates after a growth period of 6-7 days at 25 C with 14
 hrs (FL/UV) day ligh exposure. Mycelia was ground in
 liquid nitrogen prior it's storage at -80 C until RNA
 extraction. Directional cloning with 5' end of cDNA cloned
 into EcoRI site of pBluescript and 3' end of cDNA cloned
 into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
 ORIGIN

```

QY 259 ProlleValalaGlyAsnVal 265
Db 552 CYCCACATTGGCGGACTTGTG 572

RESULT 13
LOCUS BC060627
DEFINITION Mus musculus tripeptidyl peptidase II, mRNA (cDNA clone IMAGE:6844459), containing frame-shift errors.
ACCESSION BC060627
VERSION 1
KEYWORDS GI:38148645
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 4662)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, C.M., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Ioshizuka, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouford, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
2 (bases 1 to 4662)
12477932
Strausberg, R.
Direct Submission
Submitted (31-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bentso-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A., Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K., Schetz, T., Smith, C., Shir, E., Tack, D., Trout, K., Walters, J., Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILML at: http://image.llnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6678418
This clone has the following problem: frame shifted.
Location/Qualifiers
1. 4662
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"

FEATURES
source

```

```

/db xref="taxon:10090"
/clone="IMAGE:6844459"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_id="NIH_BMAP_FY0"
/lab_host="DHIOB"
/notes="vector: pYX-ASC"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 9,36e-07 Length: 4662
Score: 186.50 Matches: 101
Percent Similarity: 35.77% Conservative: 70
Best Local Similarity: 21.13% Mismatches: 154
Query Match: 8.30% Indels: 153
DB: 11 Gaps: 21

```

US-09-985-689a-1-COPY (1-434) x BC060627 (1-4662)

```

QY 10 AlaSpValAlaGlnSerSeryTyrGlyLeuTyrGlyGlnGlyGlnIleValAlaValAla 29
Db 803 GCTGAGATGCTGAATTACTCTGTGAACATTATGACGATGGAACTGCTCTCCATTGTG 862
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArgGlyLys 49
Db 863 ACCAGCGGA----- 871
QY 50 IleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGly 69
Db 872 -----GGAGCTCATGGA 883
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----GlySerThrAsnLysGlyMet 86
Db 884 ACCCATGTAGCAAGTATAGCGCGAGGCGCATTTTCCAGAGAGCGCTCAACGGAATGGATT 943
QY 87 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 99
Db 944 GCTCTGTGTGCTCAAAATTTCTATCCATTAAAGATTGGTGATACACGCTAAGCAGCTATGGA 1003
QY 100 SerGlyGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla--- 118
Db 1004 ACAGGCACAGCGCTC-----ATCAGAGCTATGATAGAGTTAATAATCATAGTGTGAT 1057
QY 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db 1058 CTTGTCAACTACAGTTATGAGAGCAACTCATTTGGCCAAATTTCTGGAGAAATT----- 1111
QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThr 155
Db 1112 -----TGTAAGTAATTATTAAGCAGTATGGAACAATACATAACAATT 1153
QY 156 IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro---Gly 174
Db 1154 TATGTTTCAAGTGTGGAATAATGTTCCATGCCCTTTCTACAGTGGTGTGTCAGAGGGA 1213
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
Db 1214 ACTACATCAGTGTGATAGTGTGAGCT----- 1243
QY 195 TyrAlaAspAsnIleAsnHisValAlaGlnPhe----- 205
Db 1244 TATGTTCCCTTGATATGATGTTTCAGAGTATTCTCAGAGAGAAAACCTGCTGCAAAAT 1303
QY 206 -----SerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 1304 CAATATACATGCTCTCTTAGAGGCCCAAGCTGATGAGGCCCTCGGTGTGAGCATGATG 1363
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241
Db 1364 GCACCAAGGAGGTGCTATTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1399
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
Db 1400 ACATTGAGGGGAGCTCAG-----CTAATGAATGGAGACATCAATGCTCTTCCCCCAATGCC 1453

```


COMMENT

Contact: Screen SE

Entomology

University of Maryland

4112 Plant Sciences Building, College Park, MD 20742, USA.

Location/Qualifiers

FEATURES

source

1. .614

/organism="Metarhizium anisopliae"

/mol_type="mRNA"

/strain="ARSEF 2575"

/db_xref="taxon:5530"

/clone="Ma#948"

/clone_lib="Metarhizium anisopliae ARSEF 2575"

/note="Vector: Unizap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional lambda vector, Unizap"

ORIGIN

Alignment Scores: 7.76e-08 Length: 614
 Pred. No.: 182.50 Matches: 67
 Score: 182.50 Conservative: 38
 Percent Similarity: 42.68%
 Best Local Similarity: 27.24% Mismatches: 88
 Query Match: 8.12% Indels: 53
 DB: 9 Gaps: 12

US-09-985-689A-1-COPY (1-434) x AJ273402 (1-614)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 11 GGTGAGGGTACTTGGCTATATATACATGACACTGGTATTGAG----- 52
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 53 GCCTCCACCCCGAGTTTGGGTGCGGCCACTTTTCTTAAGAGCTTCATCAGCGGTCAA 112
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 113 AAC---ACTGATGCCACGCCATGGGACTCACTGGCTGGTACCATT-----GGT 160
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 161 AGCAAGACCTACCGTGTGTCACAAAGGCTAAGCTCTATGGTCAAGGTTCCTTGACAA 220
 QY 101 ----GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
 DB 221 CAGGGCAGTGGTCTCTACTCCGTATCATCATGTCGTCGTCGTCGTCGTCGTCGTCGTC 280
 QY 119 TyrSerAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
 DB 281 AAGACCGCGGCTGCCCAACGCGCCATTGCTTCATGAGCCTGGGA----- 328
 QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
 DB 329 GGTGGCTACTCGGCGTCCGTCACCAAGGTGTCGTCGTCGTCGTCGTCGTCGTCGTC 388
 QY 156 IleLeuPheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
 DB 389 CTTCCGCTGCGCGTGGACGATAACCGGGATGCCCAACACCTCT---CCCGCTTCC 445
 QY 176 AlalysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly 193
 DB 446 GAGCCTCTGCTGCATCTGTGGCTCTCGGGAATAATGACAGCCGATCTTCTCTCTTC 505
 QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
 DB 506 AACTAC----- 511
 QY 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
 DB 512 GGCAGAGT---GTCGATATTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559
 QY 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253

Db 560 -----TGGATTGGT-----GGCGGCACAAACACCATCTCTGGT 592
 QY 254 ThrSerMetAlaThrPro 259
 Db 593 ACCTNCATGGCTACTGCC 610

Search completed: March 15, 2004, 22:27:57
 Job time : 2398 secs